PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/31, C07K 14/315, 16/12, C12Q 1/68

A2

(11) International Publication Number: WO 98/18931

(43) International Publication Date: 7 May 1998 (07.05.98)

(21) International Application Number: PCT/US97/19588

(22) International Filing Date: 30 October 1997 (30.10.97)

(30) Priority Data:
60/029,960 31 October 1996 (31.10.96) US

(71) Applicant *(for all designated States except US)*: HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).

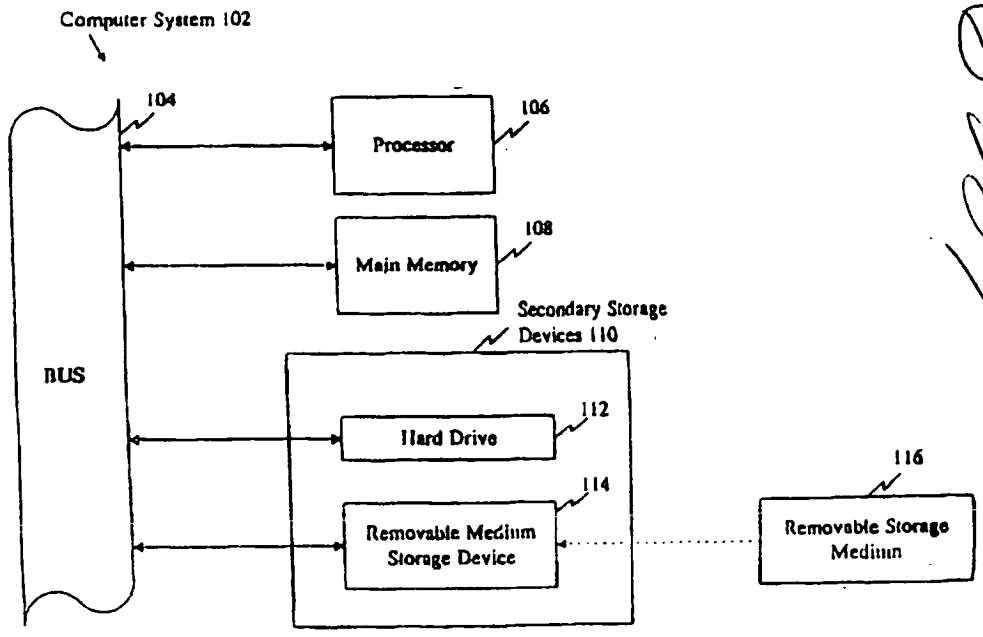
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: KUNSCH, Charles, A. [US/US]; 2398B Dunwoody Crossing, Atlanta, GA 30338 (US). CHOI, Gil, H. [KR/US]; 11429 Potomac Oaks Drive, Rockville, MD 20850 (US). DILLON, Patrick, J. [US/US]; 1055 Snipe Court, Carlsbad, CA 92009 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). BARASH, Steven, C. [US/US]; 582 College Parkway #303, Rockville, MD 20850 (US). FANNON, Michael [US/US]; 13501 Rippling Brook Drive, Silver Spring, MD 20850 (US). DOUGHERTY, Brian, A. [US/US]; 708 Meadow Field Court, Mount Airy, MD 21771 (US).

(74) Agents: BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: *STREPTOCOCCUS PNEUMONIAE* POLYNUCLEOTIDES AND SEQUENCES



(57) Abstract

The present invention provides polynucleotide sequences of the genome of *Streptococcus pneumoniae*, polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynucleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

Streptococcus pneumoniae Polynucleotides and Sequences

FIELD OF THE INVENTION

The present invention relates to the field of molecular biology. In particular, it relates to, among other things, nucleotide sequences of *Streptococcus pneumoniae*, contigs, ORFs, fragments, probes, primers and related polynucleotides thereof, peptides and polypeptides encoded by the sequences, and uses of the polynucleotides and sequences thereof, such as in fermentation, polypeptide production, assays and pharmaceutical development, among others.

BACKGROUND OF THE INVENTION

Streptococcus pneumoniae has been one of the most extensively studied microorganisms since its first isolation in 1881. It was the object of many investigations that led to important scientific discoveries. In 1928, Griffith observed that when heat-killed encapsulated pneumococci and live strains constitutively lacking any capsule were concomitantly injected into mice, the nonencapsulated could be converted into encapsulated pneumococci with the same capsular type as the heat-killed strain. Years later, the nature of this "transforming principle," or carrier of genetic information, was shown to be DNA. (Avery, O.T., *et al.*, *J. Exp. Med.*, *79:*137-157 (1944)).

In spite of the vast number of publications on *S. pneumoniae* many questions about its virulence are still unanswered, and this pathogen remains a major causative agent of serious human disease, especially community-acquired pneumonia. (Johnston, R.B., *et al.*, *Rev. Infect. Dis. 13*(Suppl. 6):S509-517 (1991)). In addition, in developing countries, the pneumococcus is responsible for the death of a large number of children under the age of 5 years from pneumococcal pneumonia. The incidence of pneumococcal disease is highest in infants under 2 years of age and in people over 60 years of age. Pneumococci are the second most frequent cause (after *Haemophilus influenzae* type b) of bacterial meningitis and otitis media in children. With the recent introduction of conjugate vaccines for *H. influenzae* type b, pneumococcal meningitis is likely to become increasingly prominent. *S. pneumoniae* is the most important etiologic agent of community-

acquired pneumonia in adults and is the second most common cause of bacterial meningitis behind *Neisseria meningitidis*.

The antibiotic generally prescribed to treat *S. pneumoniae* is benzylpenicillin, although resistance to this and to other antibiotics is found occasionally. Pneumococcal resistance to penicillin results from mutations in its penicillin-binding proteins. In uncomplicated pneumococcal pneumonia caused by a sensitive strain, treatment with penicillin is usually successful unless started too late. Erythromycin or clindamycin can be used to treat pneumonia in patients hypersensitive to penicillin, but resistant strains to these drugs exist. Broad spectrum antibiotics (e.g., the tetracyclines) may also be effective, although tetracycline-resistant strains are not rare. In spite of the availability of antibiotics, the mortality of pneumococcal bacteremia in the last four decades has remained stable between 25 and 29%. (Gillespie, S.H., *et al.*, *J. Med. Microbiol. 28*:237-248 (1989).

S. pneumoniae is carried in the upper respiratory tract by many healthy individuals. It has been suggested that attachment of pneumococci is mediated by a disaccharide receptor on fibronectin, present on human pharyngeal epithelial cells. (Anderson, B.J., *et al.*, *J. Immunol. 142*:2464-2468 (1989). The mechanisms by which pneumococci translocate from the nasopharynx to the lung, thereby causing pneumonia, or migrate to the blood, giving rise to bacteremia or septicemia, are poorly understood. (Johnston, R.B., *et al.*, *Rev. Infect. Dis. 13*(Suppl. 6):S509-517 (1991).

Various proteins have been suggested to be involved in the pathogenicity of *S. pneumoniae*, however, only a few of them have actually been confirmed as virulence factors. Pneumococci produce an IgA1 protease that might interfere with host defense at mucosal surfaces. (Kornfield, S.J., *et al.*, *Rev. Inf. Dis. 3*:521-534 (1981). *S. pneumoniae* also produces neuraminidase, an enzyme that may facilitate attachment to epithelial cells by cleaving sialic acid from the host glycolipids and gangliosides. Partially purified neuraminidase was observed to induce meningitis-like symptoms in mice; however, the reliability of this finding has been questioned because the neuraminidase preparations used were probably contaminated with cell wall products. Other pneumococcal proteins besides neuraminidase are involved in the adhesion of pneumococci to epithelial and endothelial cells. These pneumococcal proteins have as yet not been identified. Recently, Cundell *et. al.*, reported that peptide permeases can modulate

pneumococcal adherence to epithelial and endothelial cells. It was, however, unclear whether these permeases function directly as adhesions or whether they enhance adherence by modulating the expression of pneumococcal adhesions. (DeVelasco, E.A., *et al.*, *Micro. Rev.* 59:591-603 (1995). A better understanding of the virulence factors determining its pathogenicity will need to be developed to cope with the devastating effects of pneumococcal disease in humans.

Ironically, despite the prominent role of *S. pneumoniae* in the discovery of DNA, little is known about the molecular genetics of the organism. The *S. pneumoniae* genome consists of one circular, covalently closed, double-stranded DNA and a collection of so-called variable accessory elements, such as prophages, plasmids, transposons and the like. Most physical characteristics and almost all of the genes of *S. pneumoniae* are unknown. Among the few that have been identified, most have not been physically mapped or characterized in detail. Only a few genes of this organism have been sequenced. (See, for instance current versions of GENBANK and other nucleic acid databases, and references that relate to the genome of *S. pneumoniae* such as those set out elsewhere herein.)

It is clear that the etiology of diseases mediated or exacerbated by *S. pneumoniae*, infection involves the programmed expression of *S. pneumoniae* genes, and that characterizing the genes and their patterns of expression would add dramatically to our understanding of the organism and its host interactions. Knowledge of *S. pneumoniae* genes and genomic organization would improve our understanding of disease etiology and lead to improved and new ways of preventing, ameliorating, arresting and reversing diseases. Moreover, characterized genes and genomic fragments of *S. pneumoniae* would provide reagents for, among other things, detecting, characterizing and controlling *S. pneumoniae* infections. There is a need to characterize the genome of *S. pneumoniae* and for polynucleotides of this organism.

SUMMARY OF THE INVENTION

The present invention is based on the sequencing of fragments of the *Streptococcus pneumoniae* genome. The primary nucleotide sequences which were generated are provided in SEQ ID NOS:1-391.

The present invention provides the nucleotide sequence of several hundred contigs of the *Streptococcus pneumoniae* genome, which are listed in tables below and set out in the Sequence Listing submitted herewith, and representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan. In one embodiment, the present invention is provided as contiguous strings of primary sequence information corresponding to the nucleotide sequences depicted in SEQ ID NOS:1-391.

The present invention further provides nucleotide sequences which are at least 95% identical to the nucleotide sequences of SEQ ID NOS:1-391.

The nucleotide sequence of SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NOS:1-391 may be provided in a variety of mediums to facilitate its use. In one application of this embodiment, the sequences of the present invention are recorded on computer readable media. Such media includes, but is not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

The present invention further provides systems, particularly computer-based systems which contain the sequence information herein described stored in a data storage means. Such systems are designed to identify commercially important fragments of the *Streptococcus pneumoniae* genome.

Another embodiment of the present invention is directed to fragments of the *Streptococcus pneumoniae* genome having particular structural or functional attributes. Such fragments of the *Streptococcus pneumoniae* genome of the present invention include, but are not limited to, fragments which encode peptides, hereinafter referred to as open reading frames or ORFs, fragments which modulate the expression of an operably linked ORF, hereinafter referred to as expression modulating fragments or EMFs, and fragments which can be used to diagnose the

presence of *Streptococcus pneumoniae* in a sample, hereinafter referred to as diagnostic fragments or DFs.

Each of the ORFs in fragments of the *Streptococcus pneumoniae* genome disclosed in Tables 1-3, and the EMFs found 5' to the ORFs, can be used in numerous ways as polynucleotide reagents. For instance, the sequences can be used as diagnostic probes or amplification primers for detecting or determining the presence of a specific microbe in a sample, to selectively control gene expression in a host and in the production of polypeptides, such as polypeptides encoded by ORFs of the present invention, particular those polypeptides that have a pharmacological activity.

The present invention further includes recombinant constructs comprising one or more fragments of the *Streptococcus pneumoniae* genome of the present invention. The recombinant constructs of the present invention comprise vectors, such as a plasmid or viral vector, into which a fragment of the *Streptococcus pneumoniae* has been inserted.

The present invention further provides host cells containing any of the isolated fragments of the *Streptococcus pneumoniae* genome of the present invention. The host cells can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic cell, such as a yeast cell, or a procaryotic cell such as a bacterial cell.

The present invention is further directed to isolated polypeptides and proteins encoded by ORFs of the present invention. A variety of methods, well known to those of skill in the art, routinely may be utilized to obtain any of the polypeptides and proteins of the present invention. For instance, polypeptides and proteins of the present invention having relatively short, simple amino acid sequences readily can be synthesized using commercially available automated peptide synthesizers. Polypeptides and proteins of the present invention also may be purified from bacterial cells which naturally produce the protein. Yet another alternative is to purify polypeptide and proteins of the present invention from cells which have been altered to express them.

The invention further provides methods of obtaining homologs of the fragments of the *Streptococcus pneumoniae* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. Specifically, by using the nucleotide and amino acid sequences disclosed herein as

a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

The invention further provides antibodies which selectively bind polypeptides and proteins of the present invention. Such antibodies include both monoclonal and polyclonal antibodies.

The invention further provides hybridomas which produce the above-described antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

The present invention further provides methods of identifying test samples derived from cells which express one of the ORFs of the present invention, or a homolog thereof. Such methods comprise incubating a test sample with one or more of the antibodies of the present invention, or one or more of the DFs of the present invention, under conditions which allow a skilled artisan to determine if the sample contains the ORF or product produced therefrom.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the above-described assays.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the antibodies, or one of the DFs of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of bound antibodies or hybridized DFs.

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents capable of binding to a polypeptide or protein encoded by one of the ORFs of the present invention. Specifically, such agents include, as further described below, antibodies, peptides, carbohydrates, pharmaceutical agents and the like. Such methods comprise steps of: (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention; and (b) determining whether the agent binds to said protein.

The present genomic sequences of *Streptococcus pneumoniae* will be of great value to all laboratories working with this organism and for a variety of commercial purposes. Many fragments of the *Streptococcus pneumoniae* genome will be immediately identified by similarity searches against GenBank or protein databases and will be of immediate value to *Streptococcus pneumoniae* researchers

and for immediate commercial value for the production of proteins or to control gene expression.

The methodology and technology for elucidating extensive genomic sequences of bacterial and other genomes has and will greatly enhance the ability to analyze and understand chromosomal organization. In particular, sequenced contigs and genomes will provide the models for developing tools for the analysis of chromosome structure and function, including the ability to identify genes within large segments of genomic DNA, the structure, position, and spacing of regulatory elements, the identification of genes with potential industrial applications, and the ability to do comparative genomic and molecular phylogeny.

DESCRIPTION OF THE FIGURES

FIGURE 1 is a block diagram of a computer system (102) that can be used to implement computer-based systems of present invention.

FIGURE 2 is a schematic diagram depicting the data flow and computer programs used to collect, assemble, edit and annotate the contigs of the *Streptococcus pneumoniae* genome of the present invention. Both Macintosh and Unix platforms are used to handle the AB 373 and 377 sequence data files, largely as described in Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, 585, IEEE Computer Society Press, Washington D.C. (1993). Factura (AB) is a Macintosh program designed for automatic vector sequence removal and end-trimming of sequence files. The program Loadis runs on a Macintosh platform and parses the feature data extracted from the sequence files by Factura to the Unix based *Streptococcus pneumoniae* relational database. Assembly of contigs (and whole genome sequences) is accomplished by retrieving a specific set of sequence files and their associated features using Extrseq, a Unix utility for retrieving sequences from an SQL database. The resulting sequence file is processed by seq_filter to trim portions of the sequences with more than 2% ambiguous nucleotides. The sequence files were assembled using TIGR Assembler, an assembly engine designed at The Institute for Genomic Research (TIGR) for rapid and accurate assembly of thousands of sequence fragments. The collection of contigs generated by the assembly step is loaded into the database with the lassie program. Identification of open reading

frames (ORFs) is accomplished by processing contigs with zorf or GenMark. The ORFs are searched against *S. pneumoniae* sequences from GenBank and against all protein sequences using the BLASTN and BLASTP programs, described in Altschul *et al., J. Mol. Biol. 215:* 403-410 (1990)). Results of the ORF determination and similarity searching steps were loaded into the database. As described below, some results of the determination and the searches are set out in Tables 1-3.

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

The present invention is based on the sequencing of fragments of the *Streptococcus pneumoniae* genome and analysis of the sequences. The primary nucleotide sequences generated by sequencing the fragments are provided in SEQ ID NOS:1-391. (As used herein, the "primary sequence" refers to the nucleotide sequence represented by the IUPAC nomenclature system.)

In addition to the aforementioned *Streptococcus pneumoniae* polynucleotide and polynucleotide sequences, the present invention provides the nucleotide sequences of SEQ ID NOS:1-391, or representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan.

As used herein, a "representative fragment of the nucleotide sequence depicted in SEQ ID NOS:1-391" refers to any portion of the SEQ ID NOS:1-391 which is not presently represented within a publicly available database. Preferred representative fragments of the present invention are *Streptococcus pneumoniae* open reading frames (ORFs), expression modulating fragment (EMFs) and fragments which can be used to diagnose the presence of *Streptococcus pneumoniae* in sample (DFs). A non-limiting identification of preferred representative fragments is provided in Tables 1-3. As discussed in detail below, the information provided in SEQ ID NOS:1-391 and in Tables 1-3 together with routine cloning, synthesis, sequencing and assay methods will enable those skilled in the art to clone and sequence all "representative fragments" of interest, including open reading frames encoding a large variety of *Streptococcus pneumoniae* proteins.

While the presently disclosed sequences of SEQ ID NOS:1-391 are highly accurate, sequencing techniques are not perfect and, in relatively rare instances, further investigation of a fragment or sequence of the invention may reveal a

nucleotide sequence error present in a nucleotide sequence disclosed in SEQ ID NOS:1-391. However, once the present invention is made available (*i.e.*, once the information in SEQ ID NOS:1-391 and Tables 1-3 has been made available), resolving a rare sequencing error in SEQ ID NOS:1-391 will be well within the skill of the art. The present disclosure makes available sufficient sequence information to allow any of the described contigs or portions thereof to be obtained readily by straightforward application of routine techniques. Further sequencing of such polynucleotide may proceed in like manner using manual and automated sequencing methods which are employed ubiquitous in the art. Nucleotide sequence editing software is publicly available. For example, Applied Biosystem's (AB) AutoAssembler can be used as an aid during visual inspection of nucleotide sequences. By employing such routine techniques potential errors readily may be identified and the correct sequence then may be ascertained by targeting further sequencing effort, also of a routine nature, to the region containing the potential error.

Even if all of the very rare sequencing errors in SEQ ID NOS:1-391 were corrected, the resulting nucleotide sequences would still be at least 95% identical, nearly all would be at least 99% identical, and the great majority would be at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-391.

As discussed elsewhere herein, polynucleotides of the present invention readily may be obtained by routine application of well known and standard procedures for cloning and sequencing DNA. Detailed methods for obtaining libraries and for sequencing are provided below, for instance. A wide variety of *Streptococcus pneumoniae* strains that can be used to prepare *S. pneumoniae* genomic DNA for cloning and for obtaining polynucleotides of the present invention are available to the public from recognized depository institutions, such as the American Type Culture Collection (ATCC). While the present invention is enabled by the sequences and other information herein disclosed, the *S. pneumoniae* strain that provided the DNA of the present Sequence Listing, Strain 7/87 14.8.91, has been deposited in the ATCC, as a convenience to those of skill in the art. As a further convenience, a library of *S. pneumoniae* genomic DNA, derived from the same strain, also has been deposited in the ATCC. The *S. pneumoniae* strain was deposited on October 10, 1996, and was given Deposit No. 55840, and the cDNA library was deposited on October 11, 1996 and was given Deposit No. 97755. The genomic fragments in the library are 15 to 20 kb

fragments generated by partial Sau3A1 digestion and they are inserted into the BamHI site in the well-known lambda-derived vector lambda DASH II (Stratagene, La Jolla, CA). The provision of the deposits is not a waiver of any rights of the inventors or their assignees in the present subject matter.

The nucleotide sequences of the genomes from different strains of *Streptococcus pneumoniae* differ somewhat. However, the nucleotide sequences of the genomes of all *Streptococcus pneumoniae* strains will be at least 95% identical, in corresponding part, to the nucleotide sequences provided in SEQ ID NOS:1-391. Nearly all will be at least 99% identical and the great majority will be 99.9% identical.

Thus, the present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-391, in a form which can be readily used, analyzed and interpreted by the skilled artisan.

Methods for determining whether a nucleotide sequence is at least 95%, at least 99% or at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-391 are routine and readily available to the skilled artisan. For example, the well known fasta algorithm described in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA 85:* 2444 (1988) can be used to generate the percent identity of nucleotide sequences. The BLASTN program also can be used to generate an identity score of polynucleotides compared to one another.

COMPUTER RELATED EMBODIMENTS

The nucleotide sequences provided in SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide sequence of SEQ ID NOS:1-391 may be "provided" in a variety of mediums to facilitate use thereof. As used herein, provided refers to a manufacture, other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention; *i.e.*, a nucleotide sequence provided in SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide of SEQ ID NOS:1-391. Such a manufacture provides a large portion of the *Streptococcus pneumoniae* genome and parts thereof (*e.g.*, a *Streptococcus pneumoniae* open reading frame (ORF)) in a form which allows a skilled artisan to examine the manufacture using

means not directly applicable to examining the *Streptococcus pneumoniae* genome or a subset thereof as it exists in nature or in purified form.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD- ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. Likewise, it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently know methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention. A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially- available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data-processor structuring formats (*e.g.*, text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form the nucleotide sequences of SEQ ID NOS:1-

391, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a sequence of SEQ ID NOS:1-391 the present invention enables the skilled artisan routinely to access the provided sequence information for a wide variety of purposes.

The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.*, *J. Mol. Biol. 215*:403-410 (1990)) and BLAZE (Brutlag *et al.*, *Comp. Chem. 17*:203-207 (1993)) search algorithms on a Sybase system was used to identify open reading frames (ORFs) within the *Streptococcus pneumoniae* genome which contain homology to ORFs or proteins from both *Streptococcus pneumoniae* and from other organisms. Among the ORFs discussed herein are protein encoding fragments of the *Streptococcus pneumoniae* genome useful in producing commercially important proteins, such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify, among other things, commercially important fragments of the *Streptococcus pneumoniae* genome.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means.

As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage

means. Search means are used to identify fragments or regions of the present genomic sequences which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *Streptococcus pneumoniae* genomic sequences possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the

Streptococcus pneumoniae genome. In the present examples, implementing software which implement the BLAST and BLAZE algorithms, described in Altschul *et al., J. Mol. Biol. 215:* 403-410 (1990), is used to identify open reading frames within the *Streptococcus pneumoniae* genome. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill also may be employed in this regard.

Figure 1 provides a block diagram of a computer system illustrative of embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, *etc.* A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, *etc.*) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for accessing and processing the genomic sequence (such as search tools, comparing tools, *etc.*) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

BIOCHEMICAL EMBODIMENTS

Other embodiments of the present invention are directed to isolated fragments of the *Streptococcus pneumoniae* genome. The fragments of the *Streptococcus pneumoniae* genome of the present invention include, but are not limited to fragments which encode peptides and polypeptides, hereinafter open reading frames (ORFs), fragments which modulate the expression of an operably linked ORF, hereinafter expression modulating fragments (EMFs) and fragments which can be used to diagnose the presence of *Streptococcus pneumoniae* in a sample, hereinafter diagnostic fragments (DFs).

As used herein, an "isolated nucleic acid molecule" or an "isolated fragment of the *Streptococcus pneumoniae* genome" refers to a nucleic acid molecule possessing a specific nucleotide sequence which has been subjected to purification means to reduce, from the composition, the number of compounds which are normally associated with the composition. Particularly, the term refers to the nucleic acid molecules having the sequences set out in SEQ ID NOS:1-391, to representative fragments thereof as described above, to polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence thereto, also as set out above.

A variety of purification means can be used to generate the isolated fragments of the present invention. These include, but are not limited to methods which separate constituents of a solution based on charge, solubility, or size.

In one embodiment, *Streptococcus pneumoniae* DNA can be enzymatically sheared to produce fragments of 15-20 kb in length. These fragments can then be used to generate a *Streptococcus pneumoniae* library by inserting them into lambda clones as described in the Examples below. Primers flanking, for example, an ORF, such as those enumerated in Tables 1-3 can then be generated using nucleotide sequence information provided in SEQ ID NOS:1-391. Well known and routine techniques of PCR cloning then can be used to isolate the ORF from the lambda DNA library or *Streptococcus pneumoniae* genomic DNA. Thus, given the availability of SEQ ID NOS:1-391, the information in Tables 1, 2 and 3, and the information that may be obtained readily by analysis of the sequences of SEQ ID NOS:1-391 using methods set out above, those of skill will be enabled by the present disclosure to isolate any ORF-containing or other nucleic acid fragment of the present invention.

The isolated nucleic acid molecules of the present invention include, but are not limited to single stranded and double stranded DNA, and single stranded RNA.

As used herein, an "open reading frame," ORF, means a series of triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

Tables 1, 2, and 3 list ORFs in the *Streptococcus pneumoniae* genomic contigs of the present invention that were identified as putative coding regions by the GeneMark software using organism-specific second-order Markov probability transition matrices. It will be appreciated that other criteria can be used, in accordance with well known analytical methods, such as those discussed herein, to generate more inclusive, more restrictive, or more selective lists.

Table 1 sets out ORFs in the *Streptococcus pneumoniae* contigs of the present invention that over a continuous region of at least 50 bases are 95% or more identical (by BLAST analysis) to a nucleotide sequence available through GenBank in October, 1997.

Table 2 sets out ORFs in the *Streptococcus pneumoniae* contigs of the present invention that are not in Table 1 and match, with a BLASTP probability score of 0.01 or less, a polypeptide sequence available through GenBank in October, 1997.

Table 3 sets out ORFs in the *Streptococcus pneumoniae* contigs of the present invention that do not match significantly, by BLASTP analysis, a polypeptide sequence available through GenBank in October, 1997.

In each table, the first and second columns identify the ORF by, respectively, contig number and ORF number within the contig; the third column indicates the first nucleotide of the ORF (actually the first nucleotide of the stop codon immediately preceeding the ORF), counting from the 5' end of the contig strand; and the fourth column, "stop (nt)" indicates the last nucleotide of the stop codon defining the 3'end of the ORF.

In Tables 1 and 2, column five, lists the Reference for the closest matching sequence available through GenBank. These reference numbers are the databases entry numbers commonly used by those of skill in the art, who will be familiar with their denominators. Descriptions of the nomenclature are available from the National Center for Biotechnology Information. Column six in Tables 1 and 2 provides the gene name of the matching sequence; column seven provides the BLAST identity score and column eight the BLAST similarity score from the

comparison of the ORF and the homologous gene; and column nine indicates the length in nucleotides of the highest scoring segment pair identified by the BLAST identity analysis.

Each ORF described in the tables is defined by "start (nt)" (5') and "stop (nt)" (3') nucleotide position numbers. These position numbers refer to the boundaries of each ORF and provide orientation with respect to whether the forward or reverse strand is the coding strand and which reading frame the coding sequence is contained. The "start" position is the first nucleotide of the triplet encoding a stop codon just 5' to the ORF and the "stop" position is the last nucleotide of the triplet encoding the next in-frame stop codon (i.e., the stop codon at the 3' end of the ORF). Those of ordinary skill in the art appreciate that preferred fragments within each ORF described in the table include fragments of each ORF which include the entire sequence from the delineated "start" and "stop" positions excepting the first and last three nucleotides since these encode stop codons. Thus, polynucleotides set out as ORFs in the tables but lacking the three (3) 5' nucleotides and the three (3) 3' nucleotides are encompassed by the present invention. Those of skill also appreciate that particularly preferred are fragments within each ORF that are polynucleotide fragments comprising polypeptide coding sequence. As defined herein, "coding sequence" includes the fragment within an ORF beginning at the first in-frame ATG (triplet encoding methionine) and ending with the last nucleotide prior to the triplet encoding the 3' stop codon. Preferred are fragments comprising the entire coding sequence and fragments comprising the entire coding sequence, excepting the coding sequence for the N-terminal methionine. Those of skill appreciate that the N-terminal methionine is often removed during post-translational processing and that polynucleotides lacking the ATG can be used to facilitate production of N-termainal fusion proteins which may be benefical in the production or use of genetically engineered proteins. Of course, due to the degeneracy of the genetic code many polynucleotides can encode a given polypeptide. Thus, the invention further includes polynucleotides comprising a nucleotide sequence encoding a polypeptide sequence itself encoded by the coding sequence within an ORF described in Tables 1-3 herein. Further, polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence to the foregoing polynucleotides, are contemplated by the present invention.

Polypeptides encoded by polynucleotides described above and elsewhere herein are also provided by the present invention as are polypeptide comprising a an amino acid sequence at least about 95%, preferably at least 97% and even more preferably 99% identical to the amino acid sequence of a polypeptide encoded by an ORF shown in Tables 1-3. These polypeptides may or may not comprise an N-terminal methionine.

The concepts of percent identity and percent similarity of two polypeptide sequences is well understood in the art. For example, two polypeptides 10 amino acids in length which differ at three amino acid positions (*e.g.*, at positions 1, 3 and 5) are said to have a percent identity of 70%. However, the same two polypeptides would be deemed to have a percent similarity of 80% if, for example at position 5, the amino acids moieties, although not identical, were "similar" (*i.e.*, possessed similar biochemical characteristics). Many programs for analysis of nucleotide or amino acid sequence similarity, such as fasta and BLAST specifically list percent identity of a matching region as an output parameter. Thus, for instance, Tables 1 and 2 herein enumerate the percent identity of the highest scoring segment pair in each ORF and its listed relative. Further details concerning the algorithms and criteria used for homology searches are provided below and are described in the pertinent literature highlighted by the citations provided below.

It will be appreciated that other criteria can be used to generate more inclusive and more exclusive listings of the types set out in the tables. As those of skill will appreciate, narrow and broad searches both are useful. Thus, a skilled artisan can readily identify ORFs in contigs of the *Streptococcus pneumoniae* genome other than those listed in Tables 1-3, such as ORFs which are overlapping or encoded by the opposite strand of an identified ORF in addition to those ascertainable using the computer-based systems of the present invention.

As used herein, an "expression modulating fragment," EMF, means a series of nucleotide molecules which modulates the expression of an operably linked ORF or EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are fragments which induce the expression or an operably linked ORF in response to a specific regulatory factor or physiological event.

EMF sequences can be identified within the contigs of the *Streptococcus pneumoniae* genome by their proximity to the ORFs provided in Tables 1-3. An intergenic segment, or a fragment of the intergenic segment, from about 10 to 200 nucleotides in length, taken from any one of the ORFs of Tables 1-3 will modulate the expression of an operably linked ORF in a fashion similar to that found with the naturally linked ORF sequence. As used herein, an "intergenic segment" refers to fragments of the *Streptococcus pneumoniae* genome which are between two ORF(s) herein described. EMFs also can be identified using known EMFs as a target sequence or target motif in the computer-based systems of the present invention. Further, the two methods can be combined and used together.

The presence and activity of an EMF can be confirmed using an EMF trap vector. An EMF trap vector contains a cloning site linked to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic resistance or a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. As described above, a EMF will modulate the expression of an operably linked marker sequence. A more detailed discussion of various marker sequences is provided below. A sequence which is suspected as being an EMF is cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host using known procedures and the phenotype of the transformed host in examined under appropriate conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence.

As used herein, a "diagnostic fragment," DF, means a series of nucleotide molecules which selectively hybridize to *Streptococcus pneumoniae* sequences. DFs can be readily identified by identifying unique sequences within contigs of the *Streptococcus pneumoniae* genome, such as by using well-known computer analysis software, and by generating and testing probes or amplification primers

consisting of the DF sequence in an appropriate diagnostic format which determines amplification or hybridization selectivity.

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequences provided in SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence at least 95%, preferrably at least 99% and most at least preferably 99.9% identical to SEQ ID NOS:1-391, with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated. Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (*i.e.*, sequence both strands). Alternatively, error screening can be performed by sequencing corresponding polynucleotides of *Streptococcus pneumoniae* origin isolated by using part or all of the fragments in question as a probe or primer.

Preferred DFs of the present invention comprise at least about 17, preferrably at least about 20, and more preferrably at least about 50 contiguous nucleotides within an ORF set out in Tables 1-3. Most highly preferred DFs specifically hybridize to a polynucleotide containing the sequence of the ORF from which they are derived. Specific hybridization occurs even under stringent conditions defined elsewhere herein.

Each of the ORFs of the *Streptococcus pneumoniae* genome disclosed in Tables 1, 2 and 3, and the EMFs found 5' to the ORFs, can be used as polynucleotide reagents in numerous ways. For example, the sequences can be used as diagnostic probes or diagnostic amplification primers to detect the presence of a specific microbe in a sample, particularly *Streptococcus pneumoniae*. Especially preferred in this regard are ORFs such as those of Table 3, which do not match previously characterized sequences from other organisms and thus are most likely to be highly selective for *Streptococcus pneumoniae*. Also particularly preferred are ORFs that can be used to distinguish between strains of *Streptococcus pneumoniae*, particularly those that distinguish medically important strain, such as drug-resistant strains.

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Information from the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription, for triple-helix formation, or to the mRNA itself, for antisense inhibition. Both techniques have been demonstrated to be effective in model systems, and the requisite techniques are well known and involve routine procedures. Triple helix techniques are discussed in, for example, Lee *et al.*, *Nucl. Acids Res.* 6:3073 (1979); Cooney *et al.*, *Science* 241:456 (1988); and Dervan *et al.*, *Science* 251:1360 (1991). Antisense techniques in general are discussed in, for instance, Okano, *J. Neurochem.* 56:560 (1991) and *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, FL (1988)).

The present invention further provides recombinant constructs comprising one or more fragments of the *Streptococcus pneumoniae* genomic fragments and contigs of the present invention. Certain preferred recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a fragment of the *Streptococcus pneumoniae* genome has been inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. For vectors comprising the EMFs of the present invention, the vector may further comprise a marker sequence or heterologous ORF operably linked to the EMF.

Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Useful bacterial vectors include phagescript, PsiX174, pBluescript SK, pBS KS, pNH8a, pNH16a, pNH18a, pNH46a (available from Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (available from Pharmacia). Useful eukaryotic vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG

(available from Stratagene) pSVK3, pBPV, pMSG, pSVL (available from Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein- I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

The present invention further provides host cells containing any one of the isolated fragments of the *Streptococcus pneumoniae* genomic fragments and contigs of the present invention, wherein the fragment has been introduced into the host cell using known methods. The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or a procaryotic cell, such as a bacterial cell.

A polynucleotide of the present invention, such as a recombinant construct comprising an ORF of the present invention, may be introduced into the host by a variety of well established techniques that are standard in the art, such as calcium phosphate transfection, DEAE, dextran mediated transfection and electroporation, which are described in, for instance, Davis, L. *et al.*, BASIC METHODS IN MOLECULAR BIOLOGY (1986).

A host cell containing one of the fragments of the *Streptococcus pneumoniae* genomic fragments and contigs of the present invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF. The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the Genetic Code, encode an identical polypeptide sequence.

Preferred nucleic acid fragments of the present invention are the ORFs and subfragments thereof depicted in Tables 2 and 3 which encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. Such short fragments as may be obtained most readily by synthesis are useful, for example, in generating antibodies against the native polypeptide, as discussed further below.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily employ well-known methods for isolating polypeptides and proteins to isolate and purify polypeptides or proteins of the present invention produced naturally by a bacterial strain, or by other methods. Methods for isolation and purification that can be employed in this regard include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography.

The polypeptides and proteins of the present invention also can be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. Those skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, CV-1 cell, COS cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level.

"Recombinant," as used herein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (*e.g.*, yeast) expression systems. As a product, "recombinant microbial"defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g.*, *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

"Nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the polypeptides and proteins provided by this invention are assembled from fragments of the *Streptococcus pneumoniae* genome and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

Recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic regulatory elements necessary for gene expression in the host, including elements required to initiate and maintain transcription at a level sufficient for suitable expression of the desired polypeptide, including, for example, promoters and, where necessary, an enhancer and a polyadenylation signal; (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate signals to initiate translation at the beginning of the desired coding region and terminate translation at its end. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

"Recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extra chromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express

heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning*: *A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference in its entirety.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3- phosphoglycerate kinase (PGK), alpha-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, when desirable, provide amplification within the host.

Suitable prokaryotic hosts for transformation include strains of *E. coli*, *B. subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas* and *Streptomyces*. Others may, also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication

derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (available form Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (available from Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter, where it is inducible, is derepressed or induced by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period to provide for expression of the induced gene product. Thereafter cells are typically harvested, generally by centrifugation, disrupted to release expressed protein, generally by physical or chemical means, and the resulting crude extract is retained for further purification.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in Gluzman, *Cell 23:*175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Recombinant polypeptides and proteins produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The present invention further includes isolated polypeptides, proteins and nucleic acid molecules which are substantially equivalent to those herein described. As used herein, substantially equivalent can refer both to nucleic acid and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between reference and subject sequences. For purposes of the present invention, sequences having equivalent biological activity, and equivalent expression characteristics are considered substantially equivalent. For purposes of determining equivalence, truncation of the mature sequence should be disregarded.

The invention further provides methods of obtaining homologs from other strains of *Streptococcus pneumoniae*, of the fragments of the *Streptococcus pneumoniae* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. As used herein, a sequence or protein of *Streptococcus pneumoniae* is defined as a homolog of a fragment of the *Streptococcus pneumoniae* fragments or contigs or a protein encoded by one of the ORFs of the present invention, if it shares significant homology to one of the fragments of the *Streptococcus pneumoniae* genome of the present invention or a protein encoded by one of the ORFs of the present invention. Specifically, by using the sequence disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

As used herein, two nucleic acid molecules or proteins are said to "share significant homology" if the two contain regions which possess greater than 85% sequence (amino acid or nucleic acid) homology. Preferred homologs in this regard are those with more than 90% homology. Especially preferred are those with 93% or more homology. Among especially preferred homologs those with 95% or more homology are particularly preferred. Very particularly preferred among these are those with 97% and even more particularly preferred among those are homologs with 99% or more homology. The most preferred homologs among these are those with 99.9% homology or more. It will be understood that, among measures of homology, identity is particularly preferred in this regard.

Region specific primers or probes derived from the nucleotide sequence provided in SEQ ID NOS:1-391 or from a nucleotide sequence at least 95%, particularly at least 99%, especially at least 99.5% identical to a sequence of SEQ

ID NOS:1-391 can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog. Methods suitable to this aspect of the present invention are well known and have been described in great detail in many publications such as, for example, Innis *et al.*, *PCR Protocols*, Academic Press, San Diego, CA (1990)).

When using primers derived from SEQ ID NOS:1-391 or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-391, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, annealing at 50-60°C in 6X SSPC and 50% formamide, and washing at 50-65°C in 0.5X SSPC) only sequences which are greater than 75% homologous to the primer will be amplified. By employing lower stringency conditions (*e.g.*, hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences which are greater than 40-50% homologous to the primer will also be amplified.

When using DNA probes derived from SEQ ID NOS:1-391, or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-391, for colony/plaque hybridization, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, hybridizing at 50- 65°C in 5X SSPC and 50% formamide, and washing at 50- 65°C in 0.5X SSPC), sequences having regions which are greater than 90% homologous to the probe can be obtained, and that by employing lower stringency conditions (*e.g.*, hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences having regions which are greater than 35-45% homologous to the probe will be obtained.

Any organism can be used as the source for homologs of the present invention so long as the organism naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs are bacteria which are closely related to *Streptococcus pneumoniae*.

ILLUSTRATIVE USES OF COMPOSITIONS OF THE INVENTION

Each ORF provided in Tables 1 and 2 is identified with a function by homology to a known gene or polypeptide. As a result, one skilled in the art can use the polypeptides of the present invention for commercial, therapeutic and industrial purposes consistent with the type of putative identification of the

polypeptide. Such identifications permit one skilled in the art to use the *Streptococcus pneumoniae* ORFs in a manner similar to the known type of sequences for which the identification is made; for example, to ferment a particular sugar source or to produce a particular metabolite. A variety of reviews illustrative of this aspect of the invention are available, including the following reviews on the industrial use of enzymes, for example, BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY HANDBOOK, 2nd Ed., MacMillan Publications, Ltd. NY (1991) and BIOCATALYSTS IN ORGANIC SYNTHESES, Tramper *et al.*, Eds., Elsevier Science Publishers, Amsterdam, The Netherlands (1985). A variety of exemplary uses that illustrate this and similar aspects of the present invention are discussed below.

1. Biosynthetic Enzymes

Open reading frames encoding proteins involved in mediating the catalytic reactions involved in intermediary and macromolecular metabolism, the biosynthesis of small molecules, cellular processes and other functions includes enzymes involved in the degradation of the intermediary products of metabolism, enzymes involved in central intermediary metabolism, enzymes involved in respiration, both aerobic and anaerobic, enzymes involved in fermentation, enzymes involved in ATP proton motor force conversion, enzymes involved in broad regulatory function, enzymes involved in amino acid synthesis, enzymes involved in nucleotide synthesis, enzymes involved in cofactor and vitamin synthesis, can be used for industrial biosynthesis.

The various metabolic pathways present in *Streptococcus pneumoniae* can be identified based on absolute nutritional requirements as well as by examining the various enzymes identified in Table 1-3 and SEQ ID NOS:1-391.

Of particular interest are polypeptides involved in the degradation of intermediary metabolites as well as non-macromolecular metabolism. Such enzymes include amylases, glucose oxidases, and catalase.

Proteolytic enzymes are another class of commercially important enzymes. Proteolytic enzymes find use in a number of industrial processes including the processing of flax and other vegetable fibers, in the extraction, clarification and depectinization of fruit juices, in the extraction of vegetables' oil and in the maceration of fruits and vegetables to give unicellular fruits. A detailed review of the proteolytic enzymes used in the food industry is provided in Rombouts *et al.*,

Symbiosis 21:79 (1986) and Voragen *et al.* in *Biocatalysts In Agricultural Biotechnology*, Whitaker *et al.*, Eds., *American Chemical Society Symposium Series 389*:93 (1989).

The metabolism of sugars is an important aspect of the primary metabolism of *Streptococcus pneumoniae*. Enzymes involved in the degradation of sugars, such as, particularly, glucose, galactose, fructose and xylose, can be used in industrial fermentation. Some of the important sugar transforming enzymes, from a commercial viewpoint, include sugar isomerases such as glucose isomerase. Other metabolic enzymes have found commercial use such as glucose oxidases which produces ketogulonic acid (KGA). KGA is an intermediate in the commercial production of ascorbic acid using the Reichstein's procedure, as described in Krueger *et al.*, *Biotechnology* 6(A), Rhine *et al.*, Eds., Verlag Press, Weinheim, Germany (1984).

Glucose oxidase (GOD) is commercially available and has been used in purified form as well as in an immobilized form for the deoxygenation of beer. See, for instance, Hartmeir *et al.*, *Biotechnology Letters 1*:21 (1979). The most important application of GOD is the industrial scale fermentation of gluconic acid. Market for gluconic acids which are used in the detergent, textile, leather, photographic, pharmaceutical, food, feed and concrete industry, as described, for example, in Bigelis *et al.*, beginning on page 357 in GENE MANIPULATIONS AND FUNGI; Benett *et al.*, Eds., Academic Press, New York (1985). In addition to industrial applications, GOD has found applications in medicine for quantitative determination of glucose in body fluids recently in biotechnology for analyzing syrups from starch and cellulose hydrosylates. This application is described in Owusu *et al.*, *Biochem. et Biophysica. Acta. 872*:83 (1986), for instance.

The main sweetener used in the world today is sugar which comes from sugar beets and sugar cane. In the field of industrial enzymes, the glucose isomerase process shows the largest expansion in the market today. Initially, soluble enzymes were used and later immobilized enzymes were developed (Krueger *et al.*, *Biotechnology, The Textbook of Industrial Microbiology*, Sinauer Associated Incorporated, Sunderland, Massachusetts (1990)). Today, the use of glucose- produced high fructose syrups is by far the largest industrial business using immobilized enzymes. A review of the industrial use of these enzymes is provided by Jorgensen, Starch *40*:307 (1988).

Proteinases, such as alkaline serine proteinases, are used as detergent additives and thus represent one of the largest volumes of microbial enzymes used in the industrial sector. Because of their industrial importance, there is a large body of published and unpublished information regarding the use of these enzymes in industrial processes. (See Faultman *et al.*, Acid Proteases Structure Function and Biology, Tang, J., ed., Plenum Press, New York (1977) and Godfrey *et al.*, Industrial Enzymes, MacMillan Publishers, Surrey, UK (1983) and Hepner *et al.*, Report Industrial Enzymes by 1990, Hel Hepner & Associates, London (1986)).

Another class of commercially usable proteins of the present invention are the microbial lipases, described by, for instance, Macrae *et al.*, *Philosophical Transactions of the Chiral Society of London 310*:227 (1985) and Poserke, *Journal of the American Oil Chemist Society 61*:1758 (1984). A major use of lipases is in the fat and oil industry for the production of neutral glycerides using lipase catalyzed inter-esterification of readily available triglycerides. Application of lipases include the use as a detergent additive to facilitate the removal of fats from fabrics in the course of the washing procedures.

The use of enzymes, and in particular microbial enzymes, as catalyst for key steps in the synthesis of complex organic molecules is gaining popularity at a great rate. One area of great interest is the preparation of chiral intermediates. Preparation of chiral intermediates is of interest to a wide range of synthetic chemists particularly those scientists involved with the preparation of new pharmaceuticals, agrochemicals, fragrances and flavors. (See Davies *et al.*, *Recent Advances in the Generation of Chiral Intermediates Using Enzymes*, CRC Press, Boca Raton, Florida (1990)). The following reactions catalyzed by enzymes are of interest to organic chemists: hydrolysis of carboxylic acid esters, phosphate esters, amides and nitriles, esterification reactions, trans-esterification reactions, synthesis of amides, reduction of alkanones and oxoalkanates, oxidation of alcohols to carbonyl compounds, oxidation of sulfides to sulfoxides, and carbon bond forming reactions such as the aldol reaction.

When considering the use of an enzyme encoded by one of the ORFs of the present invention for biotransformation and organic synthesis it is sometimes necessary to consider the respective advantages and disadvantages of using a microorganism as opposed to an isolated enzyme. Pros and cons of using a whole cell system on the one hand or an isolated partially purified enzyme on the other

hand, has been described in detail by Bud *et al.*, Chemistry in Britain (1987), p. 127.

Amino transferases, enzymes involved in the biosynthesis and metabolism of amino acids, are useful in the catalytic production of amino acids. The advantages of using microbial based enzyme systems is that the amino transferase enzymes catalyze the stereo- selective synthesis of only L-amino acids and generally possess uniformly high catalytic rates. A description of the use of amino transferases for amino acid production is provided by Roselle-David, *Methods of Enzymology 136*:479 (1987).

Another category of useful proteins encoded by the ORFs of the present invention include enzymes involved in nucleic acid synthesis, repair, and recombination.

2. Generation of Antibodies

As described here, the proteins of the present invention, as well as homologs thereof, can be used in a variety of procedures and methods known in the art which are currently applied to other proteins. The proteins of the present invention can further be used to generate an antibody which selectively binds the protein. Such antibodies can be either monoclonal or polyclonal antibodies, as well fragments of these antibodies, and humanized forms.

The invention further provides antibodies which selectively bind to one of the proteins of the present invention and hybridomas which produce these antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A. M., *Monoclonal Antibody Technology: Laboratory Techniques In Biochemistry And Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth *et al.*, *J. Immunol. Methods 35:* 1-21 (1980), Kohler and Milstein, *Nature 256*:495-497 (1975)); the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today 4*:72 (1983), pgs. 77-96 of Cole *et al.*, in *Monoclonal Antibodies And Cancer Therapy*, Alan R. Liss, Inc. (1985)). Any animal (mouse, rabbit, *etc.*) which is known to produce antibodies can be immunized with the pseudogene polypeptide. Methods for immunization are well known in the art. Such methods

include subcutaneous or interperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection.

The protein which is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to coupling the antigen with a heterologous protein (such as globulin or galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells.

Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz *et al.*, *Exp. Cell Res. 175:*109-124 (1988)).

Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A. M., *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)).

Techniques described for the production of single chain antibodies (U. S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

For polyclonal antibodies, antibody containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures.

The present invention further provides the above- described antibodies in detectably labelled form. Antibodies can be detectably labelled through the use of radioisotopes, affinity labels (such as biotin, avidin, *etc.*), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, *etc.*) fluorescent labels (such as FITC or rhodamine, *etc.*), paramagnetic atoms, *etc.* Procedures for accomplishing such labeling are well-known in the art, for example see Sternberger *et al.*, *J. Histochem. Cytochem. 18:*315 (1970); Bayer, E. A. *et al.*, *Meth. Enzym. 62:*308

(1979); Engval, E. *et al.*, *Immunol. 109*:129 (1972); Goding, J. W., *J. Immunol. Meth. 13*:215 (1976)).

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and in situ assays to identify cells or tissues in which a fragment of the *Streptococcus pneumoniae* genome is expressed.

The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D. M. *et al.*, "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W. D. *et al.*, Meth. Enzym. 34 Academic Press, N. Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and in situ assays as well as for immunoaffinity purification of the proteins of the present invention.

3. Diagnostic Assays and Kits

The present invention further provides methods to identify the expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using one of the DFs or antibodies of the present invention.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the DFs of the present invention and assaying for binding of the DFs or antibodies to components within the test sample.

Conditions for incubating a DF or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the DF or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the DFs or antibodies of the present invention. Examples of such assays can be found in Chard, T., *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and*

Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the DFs or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound DF or antibody.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, *etc.*), and containers which contain the reagents used to detect the bound antibody or DF.

Types of detection reagents include labelled nucleic acid probes, labelled secondary antibodies, or in the alternative, if the primary antibody is labelled, the enzymatic, or antibody binding reagents which are capable of reacting with the labelled antibody. One skilled in the art will readily recognize that the disclosed DFs and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

4. Screening Assay for Binding Agents

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents which bind to a protein encoded by one of the ORFs of the present invention or to one of the fragments and the *Streptococcus pneumoniae* fragment and contigs herein described.

In general, such methods comprise steps of:

(a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention, or an isolated fragment of the *Streptococcus pneumoniae* genome; and

(b) determining whether the agent binds to said protein or said fragment.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention.

Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby *et al.*, "Application of Synthetic Peptides: Antisense Peptides," in *Synthetic Peptides, A User's Guide*, W. H. Freeman, NY (1992), pp. 289-307, and Kaspczak *et al.*, *Biochemistry 28*:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control.

One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.*, *Nucl. Acids Res.* 6:3073 (1979); Cooney *et al.*, *Science 241*:456 (1988); and Dervan *et al.*, *Science 251*:1360 (1991)) or to the mRNA itself (antisense - Okano, *J. Neurochem.* 56:560 (1991); *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, FL (1988)). Triple helix- formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides, and other DNA binding agents.

5. Pharmaceutical Compositions and Vaccines

The present invention further provides pharmaceutical agents which can be used to modulate the growth or pathogenicity of *Streptococcus pneumoniae*, or another related organism, *in vivo* or *in vitro*. As used herein, a "pharmaceutical agent" is defined as a composition of matter which can be formulated using known techniques to provide a pharmaceutical compositions. As used herein, the "pharmaceutical agents of the present invention" refers the pharmaceutical agents which are derived from the proteins encoded by the ORFs of the present invention or are agents which are identified using the herein described assays.

As used herein, a pharmaceutical agent is said to "modulate the growth pathogenicity of *Streptococcus pneumoniae* or a related organism, *in vivo* or *in vitro*," when the agent reduces the rate of growth, rate of division, or viability of the organism in question. The pharmaceutical agents of the present invention can modulate the growth or pathogenicity of an organism in many fashions, although an understanding of the underlying mechanism of action is not needed to practice the use of the pharmaceutical agents of the present invention. Some agents will modulate the growth by binding to an important protein thus blocking the biological activity of the protein, while other agents may bind to a component of the outer

surface of the organism blocking attachment or rendering the organism more prone to act the bodies nature immune system. Alternatively, the agent may comprise a protein encoded by one of the ORFs of the present invention and serve as a vaccine. The development and use of a vaccine based on outer membrane components are well known in the art.

As used herein, a "related organism" is a broad term which refers to any organism whose growth can be modulated by one of the pharmaceutical agents of the present invention. In general, such an organism will contain a homolog of the protein which is the target of the pharmaceutical agent or the protein used as a vaccine. As such, related organisms do not need to be bacterial but may be fungal or viral pathogens.

The pharmaceutical agents and compositions of the present invention may be administered in a convenient manner, such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, they are administered in an amount of at least about 1 mg/kg body weight and in most cases they will be administered in an amount not in excess of about 1 g/kg body weight per day. In most cases, the dosage is from about 0.1 mg/kg to about 10 g/kg body weight daily, taking into account the routes of administration, symptoms, *etc.*

The agents of the present invention can be used in native form or can be modified to form a chemical derivative. As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the molecule's solubility, absorption, biological half life, *etc.* The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, *etc.* Moieties capable of mediating such effects are disclosed in, among other sources, REMINGTON'S PHARMACEUTICAL SCIENCES (1980) cited elsewhere herein.

For example, such moieties may change an immunological character of the functional derivative, such as affinity for a given antibody. Such changes in immunomodulation activity are measured by the appropriate assay, such as a competitive type immunoassay. Modifications of such protein properties as redox or thermal stability, biological half-life, hydrophobicity, susceptibility to proteolytic degradation or the tendency to aggregate with carriers or into multimers also may

be effected in this way and can be assayed by methods well known to the skilled artisan.

The therapeutic effects of the agents of the present invention may be obtained by providing the agent to a patient by any suitable means (*e.g.*, inhalation, intravenously, intramuscularly, subcutaneously, enterally, or parenterally). It is preferred to administer the agent of the present invention so as to achieve an effective concentration within the blood or tissue in which the growth of the organism is to be controlled. To achieve an effective blood concentration, the preferred method is to administer the agent by injection. The administration may be by continuous infusion, or by single or multiple injections.

In providing a patient with one of the agents of the present invention, the dosage of the administered agent will vary depending upon such factors as the patient's age, weight, height, sex, general medical condition, previous medical history, *etc.* In general, it is desirable to provide the recipient with a dosage of agent which is in the range of from about 1 pg/kg to 10 mg/kg (body weight of patient), although a lower or higher dosage may be administered. The therapeutically effective dose can be lowered by using combinations of the agents of the present invention or another agent.

As used herein, two or more compounds or agents are said to be administered "in combination" with each other when either (1) the physiological effects of each compound, or (2) the serum concentrations of each compound can be measured at the same time. The composition of the present invention can be administered concurrently with, prior to, or following the administration of the other agent.

The agents of the present invention are intended to be provided to recipient subjects in an amount sufficient to decrease the rate of growth (as defined above) of the target organism.

The administration of the agent(s) of the invention may be for either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the agent(s) are provided in advance of any symptoms indicative of the organisms growth. The prophylactic administration of the agent(s) serves to prevent, attenuate, or decrease the rate of onset of any subsequent infection. When provided therapeutically, the agent(s) are provided at (or shortly after) the onset of an indication of infection. The therapeutic administration of the compound(s)

serves to attenuate the pathological symptoms of the infection and to increase the rate of recovery.

The agents of the present invention are administered to a subject, such as a mammal, or a patient, in a pharmaceutically acceptable form and in a therapeutically effective concentration. A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient patient. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

The agents of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in a mixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, *e.g.*, human serum albumin, are described, for example, in REMINGTON'S PHARMACEUTICAL SCIENCES, 16th Ed., Osol, A., Ed., Mack Publishing, Easton PA (1980). In order to form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Additional pharmaceutical methods may be employed to control the duration of action. Control release preparations may be achieved through the use of polymers to complex or absorb one or more of the agents of the present invention. The controlled delivery may be effectuated by a variety of well known techniques, including formulation with macromolecules such as, for example, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine, sulfate, adjusting the concentration of the macromolecules and the agent in the formulation, and by appropriate use of methods of incorporation, which can be manipulated to effectuate a desired time course of release. Another possible method to control the duration of action by controlled release preparations is to incorporate agents of the present invention into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization with, for example, hydroxymethylcellulose or gelatine-

microcapsules and poly(methylmethacylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions. Such techniques are disclosed in REMINGTON'S PHARMACEUTICAL SCIENCES (1980).

The invention further provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

In addition, the agents of the present invention may be employed in conjunction with other therapeutic compounds.

6. Shot-Gun Approach to Megabase DNA Sequencing

The present invention further demonstrates that a large sequence can be sequenced using a random shotgun approach. This procedure, described in detail in the examples that follow, has eliminated the up front cost of isolating and ordering overlapping or contiguous subclones prior to the start of the sequencing protocols.

Certain aspects of the present invention are described in greater detail in the examples that follow. The examples are provided by way of illustration. Other aspects and embodiments of the present invention are contemplated by the inventors, as will be clear to those of skill in the art from reading the present disclosure.

ILLUSTRATIVE EXAMPLES

LIBRARIES AND SEQUENCING

1. Shotgun Sequencing Probability Analysis

The overall strategy for a shotgun approach to whole genome sequencing follows from the Lander and Waterman (Landerman and Waterman, *Genomics* 2:231 (1988)) application of the equation for the Poisson distribution. According to this treatment, the probability, P , that any given base in a sequence of size L, in nucleotides, is not sequenced after a certain amount, n, in nucleotides, of random

sequence has been determined can be calculated by the equation $P = e^{-m}$, where m is L/n, the fold coverage. For instance, for a genome of 2.8 Mb, m=1 when 2.8 Mb of sequence has been randomly generated (1X coverage). At that point, $P_0 = e^{-1} = 0.37$. The probability that any given base has not been sequenced is the same as the probability that any region of the whole sequence L has not been determined and, therefore, is equivalent to the fraction of the whole sequence that has yet to be determined. Thus, at one-fold coverage, approximately 37% of a polynucleotide of size L, in nucleotides has not been sequenced. When 14 Mb of sequence has been generated, coverage is 5X for a 2.8 Mb and the unsequenced fraction drops to .0067 or 0.67%. 5X coverage of a 2.8 Mb sequence can be attained by sequencing approximately 17,000 random clones from both insert ends with an average sequence read length of 410 bp.

Similarly, the total gap length, G, is determined by the equation $G = Le^{-m}$, and the average gap size, g, follows the equation, g = L/n. Thus, 5X coverage leaves about 240 gaps averaging about 82 bp in size in a sequence of a polynucleotide 2.8 Mb long.

The treatment above is essentially that of Lander and Waterman, *Genomics* 2: 231 (1988).

2. Random Library Construction

In order to approximate the random model described above during actual sequencing, a nearly ideal library of cloned genomic fragments is required. The following library construction procedure was developed to achieve this end.

Streptococcus pneumoniae DNA is prepared by phenol extraction. A mixture containing 200 µg DNA in 1.0 ml of 300 mM sodium acetate, 10 mM Tris-HCl, 1 mM Na-EDTA, 50% glycerol is processed through a nebulizer (IPI Medical Products) with a stream of nitrogen adjusted to 35 Kpa for 2 minutes. The sonicated DNA is ethanol precipitated and redissolved in 500 µl TE buffer.

To create blunt-ends, a 100 µl aliquot of the resuspended DNA is digested with 5 units of BAL31 nuclease (New England BioLabs) for 10 min at 30°C in 200 µl BAL31 buffer. The digested DNA is phenol-extracted, ethanol-precipitated, redissolved in 100 µl TE buffer, and then size-fractionated by electrophoresis through a 1.0% low melting temperature agarose gel. The section containing DNA fragments 1.6-2.0 kb in size is excised from the gel, and the LGT agarose is melted and the resulting solution is extracted with phenol to separate the agarose from the

DNA. DNA is ethanol precipitated and redissolved in 20 µl of TE buffer for ligation to vector.

A two-step ligation procedure is used to produce a plasmid library with 97% inserts, of which >99% were single inserts. The first ligation mixture (50 ul) contains 2 µg of DNA fragments, 2 µg pUC18 DNA (Pharmacia) cut with SmaI and dephosphorylated with bacterial alkaline phosphatase, and 10 units of T4 ligase (GIBCO/BRL) and is incubated at 14°C for 4 hr. The ligation mixture then is phenol extracted and ethanol precipitated, and the precipitated DNA is dissolved in 20 µl TE buffer and electrophoresed on a 1.0% low melting agarose gel. Discrete bands in a ladder are visualized by ethidium bromide-staining and UV illumination and identified by size as insert (I), vector (v), v+I, v+2i, v+3i, *etc.* The portion of the gel containing v+I DNA is excised and the v+I DNA is recovered and resuspended into 20 µl TE. The v+I DNA then is blunt-ended by T4 polymerase treatment for 5 min. at 37°C in a reaction mixture (50 ul) containing the v+I linears, 500 µM each of the 4 dNTPs, and 9 units of T4 polymerase (New England BioLabs), under recommended buffer conditions. After phenol extraction and ethanol precipitation the repaired v+I linears are dissolved in 20 µl TE. The final ligation to produce circles is carried out in a 50 µl reaction containing 5 µl of v+I linears and 5 units of T4 ligase at 14°C overnight. After 10 min. at 70°C the following day, the reaction mixture is stored at -20°C.

This two-stage procedure results in a molecularly random collection of single-insert plasmid recombinants with minimal contamination from double-insert chimeras (<1%) or free vector (<3%).

Since deviation from randomness can arise from propagation the DNA in the host, *E. coli* host cells deficient in all recombination and restriction functions (A. Greener, *Strategies 3 (1)*:5 (1990)) are used to prevent rearrangements, deletions, and loss of clones by restriction. Furthermore, transformed cells are plated directly on antibiotic diffusion plates to avoid the usual broth recovery phase which allows multiplication and selection of the most rapidly growing cells.

Plating is carried out as follows. A 100 µl aliquot of Epicurian Coli SURE II Supercompetent Cells (Stratagene 200152) is thawed on ice and transferred to a chilled Falcon 2059 tube on ice. A 1.7 µl aliquot of 1.42 M beta-mercaptoethanol is added to the aliquot of cells to a final concentration of 25 mM. Cells are incubated on ice for 10 min. A 1 µl aliquot of the final ligation is added to the cells and incubated on ice for 30 min. The cells are heat pulsed for 30 sec. at 42°C and

placed back on ice for 2 min. The outgrowth period in liquid culture is eliminated from this protocol in order to minimize the preferential growth of any given transformed cell. Instead the transformation mixture is plated directly on a nutrient rich SOB plate containing a 5 ml bottom layer of SOB agar (5% SOB agar: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 1.5% Difco Agar per liter of media). The 5 ml bottom layer is supplemented with 0.4 ml of 50 mg/ml ampicillin per 100 ml SOB agar. The 15 ml top layer of SOB agar is supplemented with 1 ml X-Gal (2%), 1 ml MgCl (1 M), and 1 ml MgSO /100 ml SOB agar. The 15 ml top layer is poured just prior to plating. Our titer is approximately 100 colonies/10 µl aliquot of transformation.

All colonies are picked for template preparation regardless of size. Thus, only clones lost due to "poison" DNA or deleterious gene products are deleted from the library, resulting in a slight increase in gap number over that expected.

3. Random DNA Sequencing

High quality double stranded DNA plasmid templates are prepared using a "boiling bead" method developed in collaboration with Advanced Genetic Technology Corp. (Gaithersburg, MD) (Adams *et al.*, *Science 252*:1651 (1991); Adams *et al.*, *Nature 355*:632 (1992)). Plasmid preparation is performed in a 96-well format for all stages of DNA preparation from bacterial growth through final DNA purification. Template concentration is determined using Hoechst Dye and a Millipore Cytofluor. DNA concentrations are not adjusted, but low-yielding templates are identified where possible and not sequenced.

Templates are also prepared from two *Streptococcus pneumoniae* lambda genomic libraries. An amplified library is constructed in the vector Lambda GEM-12 (Promega) and an unamplified library is constructed in Lambda DASH II (Stratagene). In particular, for the unamplified lambda library, *Streptococcus pneumoniae* DNA (> 100 kb) is partially digested in a reaction mixture (200 ul) containing 50 µg DNA, 1X Sau3AI buffer, 20 units Sau3AI for 6 min. at 23°C. The digested DNA was phenol-extracted and electrophoresed on a 0.5% low melting agarose gel at 2V/cm for 7 hours. Fragments from 15 to 25 kb are excised and recovered in a final volume of 6 ul. One µl of fragments is used with 1 µl of DASHII vector (Stratagene) in the recommended ligation reaction. One µl of the ligation mixture is used per packaging reaction following the recommended protocol with the Gigapack II XL Packaging Extract (Stratagene, #227711). Phage

are plated directly without amplification from the packaging mixture (after dilution with 500 µl of recommended SM buffer and chloroform treatment). Yield is about $2.5x10^3$ pfu/ul. The amplified library is prepared essentially as above except the lambda GEM-12 vector is used. After packaging, about $3.5x10^4$ pfu are plated on the restrictive NM539 host. The lysate is harvested in 2 ml of SM buffer and stored frozen in 7% dimethylsulfoxide. The phage titer is approximately $1x10^9$ pfu/ml.

Liquid lysates (100 µl) are prepared from randomly selected plaques (from the unamplified library) and template is prepared by long-range PCR using T7 and T3 vector-specific primers.

Sequencing reactions are carried out on plasmid and/or PCR templates using the AB Catalyst LabStation with Applied Biosystems PRISM Ready Reaction Dye Primer Cycle Sequencing Kits for the M13 forward (M13-21) and the M13 reverse (M13RP1) primers (Adams *et al.*, *Nature 368*:474 (1994)). Dye terminator sequencing reactions are carried out on the lambda templates on a Perkin-Elmer 9600 Thermocycler using the Applied Biosystems Ready Reaction Dye Terminator Cycle Sequencing kits. T7 and SP6 primers are used to sequence the ends of the inserts from the Lambda GEM-12 library and T7 and T3 primers are used to sequence the ends of the inserts from the Lambda DASH II library. Sequencing reactions are performed by eight individuals using an average of fourteen AB 373 DNA Sequencers per day. All sequencing reactions are analyzed using the Stretch modification of the AB 373, primarily using a 34 cm well-to-read distance. The overall sequencing success rate very approximately is about 85% for M13-21 and M13RP1 sequences and 65% for dye-terminator reactions. The average usable read length is 485 bp for M13-21 sequences, 445bp for M13RP1 sequences, and 375 bp for dye-terminator reactions.

Richards *et al.*, Chapter 28 in AUTOMATED DNA SEQUENCING AND ANALYSIS, M. D. Adams, C. Fields, J. C. Venter, Eds., Academic Press, London, (1994) described the value of using sequence from both ends of sequencing templates to facilitate ordering of contigs in shotgun assembly projects of lambda and cosmid clones. We balance the desirability of both-end sequencing (including the reduced cost of lower total number of templates) against shorter read-lengths for sequencing reactions performed with the M13RP1 (reverse) primer compared to the M13-21 (forward) primer. Approximately one-half of the templates are sequenced from both ends. Random reverse sequencing reactions are

done based on successful forward sequencing reactions. Some M13RP1 sequences are obtained in a semi-directed fashion: M13-21: sequences pointing outward at the ends of contigs are chosen for M13RP1 sequencing in an effort to specifically order contigs.

4. Protocol for Automated Cycle Sequencing

The sequencing is carried out using ABI Catalyst robots and AB 373 Automated DNA Sequencers. The Catalyst robot is a publicly available sophisticated pipetting and temperature control robot which has been developed specifically for DNA sequencing reactions. The Catalyst combines pre-aliquoted templates and reaction mixes consisting of deoxy- and dideoxynucleotides, the thermostable Taq DNA polymerase, fluorescently-labelled sequencing primers, and reaction buffer. Reaction mixes and templates are combined in the wells of an aluminum 96-well thermocycling plate. Thirty consecutive cycles of linear amplification (*i.e.*., one primer synthesis) steps are performed including denaturation, annealing of primer and template, and extension; *i.e.*, DNA synthesis. A heated lid with rubber gaskets on the thermocycling plate prevents evaporation without the need for an oil overlay.

Two sequencing protocols are used: one for dye-labelled primers and a second for dye-labelled dideoxy chain terminators. The shotgun sequencing involves use of four dye-labelled sequencing primers, one for each of the four terminator nucleotide. Each dye-primer is labelled with a different fluorescent dye, permitting the four individual reactions to be combined into one lane of the 373 DNA Sequencer for electrophoresis, detection, and base-calling. ABI currently supplies pre-mixed reaction mixes in bulk packages containing all the necessary non-template reagents for sequencing. Sequencing can be done with both plasmid and PCR- generated templates with both dye-primers and dye- terminators with approximately equal fidelity, although plasmid templates generally give longer usable sequences.

Thirty-two reactions are loaded per AB373 Sequencer each day, for a total of 960 samples. Electrophoresis is run overnight following the manufacturer's protocols, and the data is collected for twelve hours. Following electrophoresis and fluorescence detection, the ABI 373 performs automatic lane tracking and base-calling. The lane-tracking is confirmed visually. Each sequence electropherogram (or fluorescence lane trace) is inspected visually and assessed for quality. Trailing

sequences of low quality are removed and the sequence itself is loaded via software to a Sybase database (archived daily to 8mm tape). Leading vector polylinker sequence is removed automatically by a software program. Average edited lengths of sequences from the standard ABI 373 are around 400 bp and depend mostly on the quality of the template used for the sequencing reaction. ABI 373 Sequencers converted to Stretch Liners provide a longer electrophoresis path prior to fluorescence detection and increase the average number of usable bases to 500-600 bp.

INFORMATICS

1. Data Management

A number of information management systems for a large-scale sequencing lab have been developed. (For review see, for instance, Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press, Washington D. C., 585 (1993)) The system used to collect and assemble the sequence data was developed using the Sybase relational database management system and was designed to automate data flow wherever possible and to reduce user error. The database stores and correlates all information collected during the entire operation from template preparation to final analysis of the genome. Because the raw output of the ABI 373 Sequencers was based on a Macintosh platform and the data management system chosen was based on a Unix platform, it was necessary to design and implement a variety of multi- user, client-server applications which allow the raw data as well as analysis results to flow seamlessly into the database with a minimum of user effort.

2. Assembly

An assembly engine (TIGR Assembler) developed for the rapid and accurate assembly of thousands of sequence fragments was employed to generate contigs. The TIGR assembler simultaneously clusters and assembles fragments of the genome. In order to obtain the speed necessary to assemble more than 10^4 fragments, the algorithm builds a hash table of 12 bp oligonucleotide subsequences to generate a list of potential sequence fragment overlaps. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Beginning with a single seed sequence fragment, TIGR Assembler extends the current contig by attempting to add the best matching

fragment based on oligonucleotide content. The contig and candidate fragment are aligned using a modified version of the Smith-Waterman algorithm which provides for optimal gapped alignments (Waterman, M. S., *Methods in Enzymology 164:*765 (1988)). The contig is extended by the fragment only if strict criteria for the quality of the match are met. The match criteria include the minimum length of overlap, the maximum length of an unmatched end, and the minimum percentage match. These criteria are automatically lowered by the algorithm in regions of minimal coverage and raised in regions with a possible repetitive element. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Fragments representing the boundaries of repetitive elements and potentially chimeric fragments are often rejected based on partial mismatches at the ends of alignments and excluded from the current contig. TIGR Assembler is designed to take advantage of clone size information coupled with sequencing from both ends of each template. It enforces the constraint that sequence fragments from two ends of the same template point toward one another in the contig and are located within a certain range of base pairs (definable for each clone based on the known clone size range for a given library).

The process resulted in 391 contigs as represented by SEQ ID NOs:1-391.

3. Identifying Genes

The predicted coding regions of the *Streptococcus pneumoniae* genome were initially defined with the program GeneMark, which finds ORFs using a probabilistic classification technique. The predicted coding region sequences were used in searches against a database of all nucleotide sequences from GenBank (October, 1997), using the BLASTN search method to identify overlaps of 50 or more nucleotides with at least a 95% identity. Those ORFs with nucleotide sequence matches are shown in Table 1. The ORFs without such matches were translated to protein sequences and compared to a non-redundant database of known proteins generated by combining the Swiss-prot, PIR and GenPept databases. ORFs that matched a database protein with BLASTP probability less than or equal to 0.01 are shown in Table 2. The table also lists assigned functions based on the closest match in the databases. ORFs that did not match protein or nucleotide sequences in the databases at these levels are shown in Table 3.

ILLUSTRATIVE APPLICATIONS

1. Production of an Antibody to a *Streptococcus pneumoniae* Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells using any one of the methods known in the art. The protein can also be produced in a recombinant prokaryotic expression system, such as *E. coli*, or can be chemically synthesized. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows.

2. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., *Nature 256*:495 (1975) or modifications of the methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., *Meth. Enzymol. 70*:419 (1980), and modified methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.*, *Basic Methods in Molecular Biology*, Elsevier, New York. Section 21-2 (1989).

3. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al.*, *J. Clin.* Endocrinol. Metab. 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. *et al.*, Chap. 19 in: *Handbook of Experimental Immunology*, Wier, D., ed, Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D. C. (1980)

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi- quantitatively or qualitatively to identify the presence of antigen in a biological sample. In addition, antibodies are useful in various animal models of pneumococcal disease as a means of evaluating the protein used to make the antibody as a potential vaccine target or as a means of evaluating the antibody as a potential immunotherapeutic or immunoprophylactic reagent.

4. Preparation of PCR Primers and Amplification of DNA

Various fragments of the *Streptococcus pneumoniae* genome, such as those of Tables 1-3 and SEQ ID NOS:1-391 can be used, in accordance with the present invention, to prepare PCR primers for a variety of uses. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. The PCR primers and amplified DNA of this Example find use in the Examples that follow.

5. Gene expression from DNA Sequences Corresponding to ORFs

A fragment of the *Streptococcus pneumoniae* genome provided in Tables 1-3 is introduced into an expression vector using conventional technology. Techniques to transfer cloned sequences into expression vectors that direct protein translation in mammalian, yeast, insect or bacterial expression systems are well known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield *et al.*, U. S. Patent No. 5,082,767, incorporated herein by this reference.

The following is provided as one exemplary method to generate polypeptide(s) from cloned ORFs of the *Streptococcus pneumoniae* genome fragment. Bacterial ORFs generally lack a poly A addition signal. The addition signal sequence can be added to the construct by, for example, splicing out the poly A addition sequence from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene) for use in eukaryotic expression systems. pXT1 contains the LTRs and a portion of the gag gene of Moloney Murine Leukemia Virus. The positions of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The *Streptococcus pneumoniae* DNA is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the *Streptococcus pneumoniae* DNA and containing restriction endonuclease sequences for PstI incorporated into the 5' primer and BglII at the 5' end of the corresponding *Streptococcus pneumoniae* DNA 3' primer, taking care to ensure that the *Streptococcus pneumoniae* DNA is positioned such that its followed with the poly A addition sequence. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with BglII, purified and ligated to pXT1, now containing a poly A addition sequence and digested BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600 ug/ml G418 (Sigma, St. Louis, Missouri). The protein is preferably released into the supernatant. However if the protein has membrane binding domains, the protein may additionally be retained within the cell or expression may be restricted to the cell surface. Since it may be necessary to purify and locate the transfected product, synthetic 15-mer peptides synthesized from the predicted *Streptococcus pneumoniae* DNA sequence are injected into mice to generate antibody to the polypeptide encoded by the *Streptococcus pneumoniae* DNA.

Alternatively and if antibody production ~~is not~~ possible, the *Streptococcus pneumoniae* DNA sequence is additionally incorporated into eukaryotic expression vectors and expressed as, for example, a globin fusion. Antibody to the globin moiety then is used to purify the chimeric protein. Corresponding protease cleavage sites are engineered between the globin moiety and the polypeptide encoded by the *Streptococcus pneumoniae* DNA so that the latter may be freed from the formed by simple protease digestion. One useful expression vector for generating globin chimerics is pSG5 (Stratagene). This vector encodes a rabbit globin. Intron II of the rabbit globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, cited elsewhere herein, and many of the methods are available from the technical assistance representatives from Stratagene, Life Technologies, Inc., or Promega. Polypeptides of the invention also may be produced using *in vitro* translation systems such as *in vitro* ExpressTM Translation Kit (Stratagene).

While the present invention has been described in some detail for purposes of clarity and understanding, one skilled in the art will appreciate that various changes in form and detail can be made without departing from the true scope of the invention.

All patents, patent applications and publications referred to above are hereby incorporated by reference.

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
1	1	437	1003	gb|U41735|	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	92	200	567
2	5	6169	5720	gb|U04047|	Streptococcus pneumoniae SSZ dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	96	450	450
2	6	6592	6167	emb|Z83335|SPZ8	S.pneumoniae dexB, capI[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	98	426	426
3	11	9770	9147	emb|Z83335|SPZ8	S.pneumoniae dexB, capI[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	94	624	624
3	12	10489	9671	emb|Z83335|SPZ8	S pneumoniae dexB, capI[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	91	819	819
3	13	11546	12019	gb|U43526|	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	474	474
3	14	12017	13375	gb|U43526|	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	1359	1359
3	15	13421	14338	gb|U43526|	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	918	918
3	16	14329	15171	gb|U43526|	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	843	843
3	17	15132	17282	gb|U43526|	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	2151	2151
3	18	17267	18397	gb|U43526|	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	1069	1131
4	1	46	1188	emb|Y11463|SPDN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	99	1143	1143
4	2	1198	2529	emb|Y11463|SPDN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	99	876	1332
5	7	11297	11473	gb|U41735|	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	82	175	177
6	7	7125	7364	emb|Z77726|SPIS	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	93	238	240
6	8	7322	7570	emb|Z77725|SPIS	S.pneumoniae DNA for insertion sequence IS1381 (966 bp)	95	160	249
6	9	7533	7985	emb|Z77725|SPIS	S.pneumoniae DNA for insertion sequence IS1381 (966 bp)	99	453	453
6	23	20197	19733	emb|Z83335|SPZ8	S pneumoniae dexB, capI[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	96	465	465
7	10	8305	7682	emb|Z83335|SPZ8	S pneumoniae dexB, capI[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	95	624	624

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
7	11	9024	8206	emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	95	819	819
10	13	9304	8078	gb|L29323|	Streptococcus pneumoniae methyl transferase (mtr) gene cluster, complete cds	93	513	1227
11	2	548	919	emb|Z79691|SOOR	S.pneumoniae yorf(A,B,C,D,E), ftsL, pbpX and regR genes	99	316	372
11	3	892	1980	emb|Z79691|SOOR	S.pneumoniae yorf(A,B,C,D,E), ftsL, pbpX and regR genes	99	1089	1089
11	5	3040	3477	emb|Z79691|SOOR	S.pneumoniae yorf(A,B,C,D,E), ftsL, pbpX and regR genes	99	259	438
11	6	3480	3247	emb|Z79691|SOOR	S.pneumoniae yorf(A,B,C,D,E), ftsL, pbpX and regR genes	99	234	234
11	7	3601	4557	emb|Z79691|SOOR	S.pneumoniae yorf(A,B,C,D,E), ftsL, pbpX and regR genes	98	957	957
11	8	4506	4886	emb|Z79691|SOOR	S.pneumoniae yorf(A,B,C,D,E), ftsL, pbpX and regR genes	99	381	381
11	9	4884	7142	emb|X16367|SPPB	Streptococcus pneumoniae pbpX gene for penicillin binding protein 2X	99	2259	2259
11	10	7132	8124	emb|X16367|SPPB	Streptococcus pneumoniae pbpX gene for penicillin binding protein 2X	98	70	993
13	1	53	1126	gb|M31296|	S.pneumoniae recP gene, complete cds	99	437	1074
14	3	1837	2148	emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	87	96	312
14	4	2518	2108	gb|M16180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	98	411	411
15	9	8942	8511	gb|U09239|	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cps19fABCDEFGHIJKLMNO) genes, complete cds, and aliA gene, partial cds	89	340	432
17	7	3910	3458	emb|Z77726|SPIS	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	98	453	453
17	8	4304	3873	emb|Z77727|SPIS	S.pneumoniae DNA for insertion sequence IS1318 (823 bp)	96	382	432
19	1	41	529	emb|X94909|SPIG	S.pneumoniae iga gene	75	368	489
19	2	554	757	gb|L07752|	Streptococcus pneumoniae attachment site (attB), DNA sequence	99	167	204
19	3	946	1827	gb|L07752|	Streptococcus pneumoniae attachment site (attB), DNA sequence	94	100	882
20	1	937	182	gb|U33315|	Streptococcus pneumoniae orfL gene, partial cds, competence stimulating peptide precursor (comC), histidine protein kinase (comD) and response regulator (comE) genes, complete cds, tRNA-Arg and tRNA-Gln genes	99	756	756
20	2	2271	931	gb|U33315|	Streptococcus pneumoniae orfL gene, partial cds, competence stimulating peptide precursor (comC), histidine protein kinase (comD) and response regulator (comE) genes, complete cds, tRNA-Arg and tRNA-Gln genes	98	1341	1341

TABLE 1 S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
20	3	3175	2684	gb|U76218|	Streptococcus pneumoniae competence stimulating peptide precursor ComC (comC), histidine kinase homolog ComD (comD), and response regulator homolog ComE (comE) genes, complete cds	99	492	492
20	4	3322	4527	gb|AF000658|	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtra), SPSpoJ (spspoJ), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnan) genes, complete cds	99	1206	1206
20	5	4573	5343	gb|AF000658|	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtra), SPSpoJ (spspoJ), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnan) genes, complete cds	99	771	771
20	6	5532	6917	gb|AF000658|	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtra), SPSpoJ (spspoJ), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnan) genes, complete cds	99	1386	1386
20	7	6995	8212	gb|AF000658|	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtra), SPSpoJ (spspoJ), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnan) genes, complete cds	99	1218	1218
20	8	8214	8471	gb|AF000658|	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtra), SPSpoJ (spspoJ), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnan) genes, complete cds	98	258	258
20	9	8534	9670	gb|AF000658|	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtra), SPSpoJ (spspoJ), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnan) genes, complete cds	99	134	1137
22	14	11887	12267	emb|Z77726|SPIS	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	99	226	381
22	15	12708	12256	emb|Z77727|SPIS	S.pneumoniae DNA for insertion sequence IS1318 (823 bp)	97	353	453
22	16	13165	12662	emb|Z77726|SPIS	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	98	504	504
22	23	18398	18910	emb|Z86112|SP28	S.pneumoniae genes encoding galacturonosyl transferase and transposase and insertion sequence IS1515	95	463	513
22	24	18829	19299	emb|Z86112|SP28	S.pneumoniae genes encoding galacturonosyl transferase and transposase and insertion sequence IS1515	99	443	471
23	5	5624	4203	emb|X52474|SPPL	S.pneumoniae ply gene for pneumolysin	99	1422	1422
23	6	6063	5629	gb|M17717|	S.pneumoniae pneumolysin gene, complete cds	98	197	435
26	1	5500	2	emb|X94909|SPIG	S.pneumoniae iga gene	87	3487	5499
26	2	5823	5584	gb|U47687|	Streptococcus pneumoniae immunoglobulin A1 protease (iga) gene, complete cds	99	151	240
26	3	6878	5685	gb|U47687|	Streptococcus pneumoniae immunoglobulin A1 protease (iga) gene, complete cds	100	50	1194

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent Ident	HSP nt length	ORF nt length
26	8	14498	14854	emb|Z83335|SPZ8	S.pneumoniae dexB, capI[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	99	338	357
26	9	14763	14924	emb|Z83335|SPZ8	S.pneumoniae dexB, capI[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	100	94	162
26	10	14922	15173	gb|U04047|	Streptococcus pneumoniae SSZ dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	242	252
28	1	80	505	emb|Z83335|SPZ8	S.pneumoniae dexB, capI[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	99	426	426
28	2	503	952	gb|U04047|	Streptococcus pneumoniae SSZ dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	450	450
28	3	780	1298	gb|U04047|	Streptococcus pneumoniae SSZ dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	96	181	519
34	1	207	1523	gb|L08611|	Streptococcus pneumoniae maltose/maltodextrin uptake (malX) and two maltodextrin permease (malC and malD) genes, complete cds	99	1317	1317
34	2	1477	2367	gb|L08611|	Streptococcus pneumoniae maltose/maltodextrin uptake (malX) and two maltodextrin permease (malC and malD) genes, complete cds	96	795	891
34	3	2593	3420	gb|L21856|	Streptococcus pneumoniae malA gene, complete cds; malR gene, complete cds	96	446	828
34	4	2790	2647	gb|L21856|	Streptococcus pneumoniae malA gene, complete cds; malR gene, complete cds	98	137	144
34	5	3418	4416	gb|L21856|	Streptococcus pneumoniae malA gene, complete cds; malR gene, complete cds	96	999	999
34	9	7764	7507	gb|U41735|	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	93	201	258
34	16	10562	10257	emb|X63602|SPBO	S.pneumoniae mmsA-Box	92	238	306
35	4	1176	1439	emb|Z83335|SPZ8	S pneumoniae dexB, capI[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	87	248	264
35	5	1458	1961	gb|U09239|	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cps19fABCDEFGHIJKLMNO) genes, complete cds, and aliA gene, partial cds	98	264	504
35	17	16172	15477	emb|X85787|SPCP	S.pneumoniae dexB, cps14A, cps14B, cps14C, cps14D, cps14E, cps14F, cps14G, cps14H, cps14I, cps14J, cps14K, cps14L, tasA genes	97	696	696
35	18	16961	16170	emb|Z83335|SPZ8	S.pneumoniae dexB, capI[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	86	792	792
35	19	17620	16871	gb|U09239|	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cps19fABCDEFGHIJKLMNO) genes, complete cds, and aliA gene, partial cds	83	750	750

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
35	20	19061	17604	emb|X85787|SPCP	S.pneumoniae dexB, cps14A, cps14B, cps14C, cps14D, cps14E, cps14F, cps14G, cps14H, cps14I, cps14J, cps14K, cps14L, tasA genes	94	1458	1458
36	19	18960	18352	gb|U40786|	Streptococcus pneumoniae surface antigen A variant precursor (psaA) and 18 kDa protein genes, complete cds, and ORF1 gene, partial cds	99	609	609
36	20	19934	18966	gb|U53509|	Streptococcus pneumoniae surface adhesin A precursor (psaA) gene, complete cds	99	969	969
37	1	2743	179	emb|Z67739|SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	2565	2565
37	2	2985	2824	emb|Z67739|SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	100	162	162
37	3	5034	3070	emb|Z67739|SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	1965	1965
37	4	5134	5790	emb|Z67739|SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	657	657
37	5	6171	5833	emb|Z67739|SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	96	339	339
38	19	12969	13268	gb|M28679|	S.pneumoniae promoter region DNA	100	64	300
39	2	1256	2137	gb|U41735|	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	99	882	882
39	3	2405	3370	gb|U41735|	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	99	966	966
40	9	5253	7208	gb|M29686|	S.pneumoniae mismatch repair (hexB) gene, complete cds	99	1956	1956
41	1	3	1037	emb|Z17307|SPRE	S.pneumoniae recA gene encoding RecA	99	1027	1035
41	2	1328	2713	emb|Z34303|SPCI	Streptococcus pneumoniae cin operon encoding the cinA, recA, dinF, lytA genes, and downstream sequences	99	1386	1386
41	3	3083	4045	gb|M13812|	S.pneumoniae autolysin (lytA) gene, complete cds	99	963	963
41	4	3272	3096	gb|M13812|	S.pneumoniae autolysin (lytA) gene, complete cds	100	177	177
41	5	3603	3860	gb|M13812|	S.pneumoniae autolysin (lytA) gene, complete cds	100	258	258
41	6	4755	5162	gb|L36660|	Streptococcus pneumoniae ORF, complete cds	98	408	408
41	7	5270	5716	gb|L36660|	Streptococcus pneumoniae ORF, complete cds	98	447	447
41	8	6112	6918	gb|L36660|	Streptococcus pneumoniae ORF, complete cds	98	431	807
41	9	6916	7119	gb|L36660|	Streptococcus pneumoniae ORF, complete cds	100	204	204
41	10	7082	7660	gb|L36660|	Streptococcus pneumoniae ORF, complete cds	97	552	579
41	11	7680	7979	gb|L36660|	Streptococcus pneumoniae ORF, complete cds	98	81	300
41	12	9169	8717	emb|Z77727|SPIS	S.pneumoniae DNA for insertion sequence IS1318 (823 bp)	97	353	453

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
41	13	9533	9132	emb|Z77725|SPIS	S.pneumoniae DNA for insertion sequence IS1381 (966 bp)	95	160	402
41	14	9669	9475	emb|Z82001|SPZ8	S.pneumoniae pcpA gene and open reading frames	100	189	195
44	5	7190	7555	emb|Z82001|SPZ8	S.pneumoniae pcpA gene and open reading frames	99	366	366
44	6	8059	7607	emb|Z77726|SPIS	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	97	453	453
44	7	8423	8022	emb|Z77725|SPIS	S.pneumoniae DNA for insertion sequence IS1381 (966 bp)	95	160	402
44	8	8559	8365	emb|Z82001|SPZ8	S.pneumoniae pcpA gene and open reading frames	100	189	195
48	9	6480	4687	gb|L39074|	Streptococcus pneumoniae pyruvate oxidase (spxB) gene, complete cds	99	1794	1794
49	2	231	2603	gb|L20561|	Streptococcus pneumoniae Exp7 gene, partial cds	100	216	2373
53	6	2407	2156	gb|U04047|	Streptococcus pneumoniae SSZ dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	242	252
53	7	2566	2405	emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	100	94	162
53	8	2833	2475	emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	99	338	357
54	13	12409	11105	emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	67	591	1305
55	22	20488	19949	emb|Z84379|HS28	S.pneumoniae dfr gene (isolate 92)	99	540	540
61	11	11864	9900	emb|Z16082|PNAL	Streptococcus pneumoniae aliB gene	98	1965	1965
63	1	3	239	gb|M18729|	S.pneumoniae mismatch repair protein (hexA) gene, complete cds	100	237	237
63	2	233	2611	gb|M18729|	S.pneumoniae mismatch repair protein (hexA) gene, complete cds	99	2330	2379
63	3	2557	2823	gb|M18729|	S pneumoniae mismatch repair protein (hexA) gene, complete cds	99	266	267
63	4	2958	4664	gb|M18729|	S.pneumoniae mismatch repair protein (hexA) gene, complete cds	95	69	1707
67	6	3770	3399	gb|L20670|	Streptococcus pneumoniae hyaluronidase gene, complete cds	96	372	372
67	7	7161	4171	gb|L20670|	Streptococcus pneumoniae hyaluronidase gene, complete cds	99	2938	2991
70	1	1	702	gb|M14340|	S.pneumoniae DpnI gene region encoding dpnC and dpnD, complete cds	100	693	702
70	2	678	1160	gb|M14340|	S.pneumoniae DpnI gene region encoding dpnC and dpnD, complete cds	100	483	483
70	3	2490	1210	gb|M14339|	S.pneumoniae DpnII gene region encoding dpnM, dpnA, dpnB, complete cds	98	462	1281
70	7	4230	4424	gb|J04234|	S.pneumoniae exodeoxyribonuclease (exoA) gene, complete cds	99	147	195
70	8	5197	4316	gb|J04234|	S.pneumoniae exodeoxyribonuclease (exoA) gene, complete cds	99	881	882

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
70	13	8108	9874	gb|L20562|	Streptococcus pneumoniae Exp8 gene, partial cds	93	234	1767
71	22	27964	28341	emb|X63602|SPBO	S.pneumoniae mmsA-Box	93	233	378
72	5	4607	3552	emb|Z26850|SPAT	S.pneumoniae (H222) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	97	102	1056
73	1	471	133	emb|X63602|SPBO	S.pneumoniae mmsA-Box	91	193	339
73	3	3658	977	gb|J04479|	S.pneumoniae DNA polymerase I (polA) gene, complete cds	99	2682	2682
73	8	4864	5379	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	98	318	516
77	3	2622	1999	emb|Z83335|SP28	S.pneumoniae dexB, capI[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	95	624	624
77	4	3341	2523	emb|Z83335|SP28	S.pneumoniae dexB, capI[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	91	819	819
78	1	341	3	emb|X77249|SPR6	S.pneumoniae (R6) ciaR/ciaH genes	99	339	339
78	2	1095	325	emb|X77249|SPR6	S.pneumoniae (R6) ciaR/ciaH genes	99	771	771
82	10	11436	10816	gb|U90721|	Streptococcus pneumoniae signal peptidase I (spi) gene, complete cds	97	621	621
82	11	12402	11434	gb|U93576|	Streptococcus pneumoniae ribonuclease HII (rnhB) gene, complete cds	98	953	969
82	12	12381	12704	gb|U93576|	Streptococcus pneumoniae ribonuclease HII (rnhB) gene, complete cds	100	51	324
83	8	3212	3550	emb|Z77727|SPIS	S.pneumoniae DNA for insertion sequence IS1318 (823 bp)	97	290	339
83	10	4662	6851	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	2190	2190
83	11	6849	8213	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	1365	1365
83	12	8236	9090	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	855	855
83	13	9283	13017	gb|L15190|	Streptococcus pneumoniae SAICAR synthetase (purC) gene, complete cds	100	107	3735
83	23	22147	23313	gb|L36923|	Streptococcus pneumoniae beta-N-acetylhexosaminidase (strH) gene, complete cds	98	218	1167
83	24	23268	23450	gb|L36923|	Streptococcus pneumoniae beta-N-acetylhexosaminidase (strH) gene, complete cds	98	172	183
83	25	27527	23505	gb|L36923|	Streptococcus pneumoniae beta-N-acetylhexosaminidase (strH) gene, complete cds	99	3826	4023

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
83	26	28472	27771	gb|L36923|	Streptococcus pneumoniae beta-N-acetylhexosaminidase (strH) gene, complete cds	99	416	702
84	4	4554	6173	emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	98	697	1620
87	6	5951	5316	emb|Z77725|SPIS	S.pneumoniae DNA for insertion sequence IS1381 (966 bp)	96	439	636
88	5	2957	3511	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	94	555	555
88	6	3466	4269	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	94	804	804
89	13	9878	10093	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	97	211	216
89	14	10062	10412	emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	97	335	351
93	10	5303	4941	emb|X63602|SPBO	S.pneumoniae mmsA-Box	89	237	363
97	4	1708	1520	gb|U41735|	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	91	140	189
99	1	89	700	emb|Z83335|SPZ8	S pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	93	592	612
99	2	1773	775	emb|X17337|SPAM	Streptococcus pneumoniae ami locus conferring aminopterin resistance	99	998	999
99	3	2794	1712	emb|X17337|SPAM	Streptococcus pneumoniae ami locus conferring aminopterin resistance	99	1083	1083
99	4	3732	2788	emb|X17337|SPAM	Streptococcus pneumoniae ami locus conferring aminopterin resistance	100	945	945
99	5	5249	3714	emb|X17337|SPAM	Streptococcus pneumoniae ami locus conferring aminopterin resistance	100	1536	1536
99	6	7262	5277	emb|X17337|SPAM	Streptococcus pneumoniae ami locus conferring aminopterin resistance	99	1986	1986
101	1	216	1538	emb|X54225|SPEN	S.pneumoniae epuA and endA genes for 7 kDa protein and membrane endonuclease	99	146	1323
101	2	1492	1719	emb|X54225|SPEN	S.pneumoniae epuA and endA genes for 7 kDa protein and membrane endonuclease	99	228	228
101	3	1694	1855	emb|X54225|SPEN	S.pneumoniae epuA and endA genes for 7 kDa protein and membrane endonuclease	100	162	162
101	4	1701	2582	emb|X54225|SPEN	S.pneumoniae epuA and endA genes for 7 kDa protein and membrane endonuclease	100	882	882
103	7	5556	5041	emb|Z95914|SPZ9	Streptococcus pneumoniae sodA gene	100	396	516
104	2	1347	1556	emb|Z77727|SPIS	S pneumoniae DNA for insertion sequence IS1318 (823 bp)	83	206	210

TABLE 1

S pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
105	5	5381	5028	emb|Z67739|SPPA	S pneumoniae parC, parE and transposase genes and unknown orf	98	353	354
105	6	6089	5379	emb|Z67739|SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	98	84	711
107	4	2785	1880	emb|X16022|SPPE	S.pneumoniae penA gene	98	72	906
107	5	2913	4988	emb|X16022|SPPE	S.pneumoniae penA gene	99	1692	2076
107	6	4981	5595	emb|X13136|SPPE	Streptococcus pneumoniae penA gene for penicillin binding protein 2B lacking N-term. (penicillin resistant strain)	91	107	615
108	9	9068	8718	emb|Z67739|SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	95	342	351
108	12	11308	10922	emb|Z67739|SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	199	387
109	3	2768	2241	emb|Z77725|SPIS	S.pneumoniae DNA for insertion sequence IS1381 (966 bp)	96	61	528
109	4	2688	2855	emb|Z77726|SPIS	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	96	148	168
109	5	2862	3269	emb|Z77727|SPIS	S.pneumoniae DNA for insertion sequence IS1318 (823 bp)	97	353	408
109	6	5320	3584	gb|M18729|	S.pneumoniae mismatch repair protein (hexA) gene, complete cds	100	371	1737
113	1	431	3	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	95	429	429
113	10	9788	8532	emb|X99400|SPDA	S.pneumoniae dacA gene and ORF	99	1257	1257
113	11	9870	10985	emb|X99400|SPDA	S.pneumoniae dacA gene and ORF	99	1116	1116
114	3	2530	2030	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	95	481	501
115	11	11303	10932	gb|U04047|	Streptococcus pneumoniae SSZ dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	372	372
117	1	897	3302	emb|X72967|SPNA	S.pneumoniae nanA gene	99	2402	2406
117	2	3277	3831	emb|X72967|SPNA	S.pneumoniae nanA gene	99	237	555
117	3	4327	3899	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	98	429	429
121	2	1369	1941	gb|U72720|	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	99	202	573
121	3	2412	4251	gb|U72720|	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	99	1842	1842
122	8	5066	5587	gb|U04047|	Streptococcus pneumoniae SSZ dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	64	451	522

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
125	1	1811	189	|gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	92	99	1623
128	15	12496	11204	|emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	91	705	1293
134	1	1	492	|emb|Y10818|SPY1	S.pneumoniae spsA gene	99	203	492
134	2	556	2652	|gb|AF019904|	Streptococcus pneumoniae choline binding protein A (cbpA) gene, partial cds	86	685	2097
134	3	1160	837	|emb|Y10818|SPY1	S.pneumoniae spsA gene	86	324	324
134	4	3952	2882	|gb|AF019904|	Streptococcus pneumoniae choline binding protein A (cbpA) gene, partial cds	98	215	1071
134	8	7992	9848	|gb|U12567|	Streptococcus pneumoniae PI3 glycerol-3-phosphate dehydrogenase (glpD) gene, partial cds, and glycerol uptake facilitator (glpF) and ORF3 genes, complete cds	99	285	1857
134	9	9846	10622	|gb|U12567|	Streptococcus pneumoniae PI3 glycerol-3-phosphate dehydrogenase (glpD) gene, partial cds, and glycerol uptake facilitator (glpF) and ORF3 genes, complete cds	99	570	777
134	10	10805	11122	|gb|U12567|	Streptococcus pneumoniae PI3 glycerol-3-phosphate dehydrogenase (glpD) gene, partial cds, and glycerol uptake facilitator (glpF) and ORF3 genes, complete cds	100	318	318
137	13	7970	8443	|gb|U09239|	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cps19fABCDEFGH1JKLMNO) genes, complete cds, and aliA gene, partial cds	90	420	474
137	14	8590	8775	|emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	94	174	186
137	15	8773	8967	|emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	98	195	195
137	16	9223	9687	|emb|Z77726|SPIS	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	96	446	465
137	17	9641	10051	|emb|Z77727|SPIS	S.pneumoniae DNA for insertion sequence IS1318 (823 bp)	96	293	411
139	10	12998	12702	|emb|X63602|SPBO	S.pneumoniae mmsA-Box	90	234	297
141	8	7805	8938	|emb|Z49988|SPMM	Streptococcus pneumoniae mmsA gene	99	338	1134
141	9	8936	10972	|emb|Z49988|SPMM	Streptococcus pneumoniae mmsA gene	99	2037	2037
141	10	11472	12467	|emb|Z49988|SPMM	Streptococcus pneumoniae mmsA gene	100	76	996
142	2	257	814	|gb|M80215|	Streptococcus pneumoniae uvs402 protein gene, complete cds	98	174	558
142	3	787	957	|gb|M80215|	Streptococcus pneumoniae uvs402 protein gene, complete cds	100	142	171
142	4	980	3022	|gb|M80215|	Streptococcus pneumoniae uvs402 protein gene, complete cds	95	1997	2043

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
142	5	3020	3595	gb|M80215|	Streptococcus pneumoniae uvs402 protein gene, complete cds	100	153	576
145	1	1	219	emb|Z35135|SPAL	S.pneumoniae allA gene for amiA-like gene A	97	185	219
145	2	171	1994	gb|L20556|	Streptococcus pneumoniae plpA gene, partial cds	99	1811	1824
145	3	2287	7599	emb|Z47210|SPDE	S.pneumoniae dexB, cap3A, cap3B and cap3C genes and orfs	99	1052	5313
145	4	9934	7766	gb|M90527|	Streptococcus pneumoniae penicillin-binding protein (ponA) gene, complete cds	99	2169	2169
145	5	10488	9922	gb|M90527|	Streptococcus pneumoniae penicillin-binding protein (ponA) gene, complete cds	99	512	567
146	1	159	4	emb|Z82002|SPZ8	S.pneumoniae pcpB and pcpC genes	98	156	156
146	2	344	90	emb|Z82002|SPZ8	S.pneumoniae pcpB and pcpC genes	98	255	255
146	16	11795	10794	emb|Z82002|SPZ8	S.pneumoniae pcpB and pcpC genes	85	276	1002
147	11	10678	10202	emb|Z21702|SPUN	S.pneumoniae ung gene and mutX genes encoding uracil-DNA glycosylase and 8-oxodGTP nucleoside triphosphatase	98	477	477
147	12	11338	10676	emb|Z21702|SPUN	S.pneumoniae ung gene and mutX genes encoding uracil-DNA glycosylase and 8-oxodGTP nucleoside triphosphatase	99	663	663
148	12	9009	8815	gb|U41735|	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	90	180	195
156	4	1154	1402	emb|X63602|SPBO	S pneumoniae mmsA-Box	94	185	249
159	13	9048	8521	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	98	526	528
160	1	1	147	emb|Z26851|SPAT	S.pneumoniae (R6) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	100	142	147
160	2	179	898	emb|Z26851|SPAT	S.pneumoniae (R6) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	99	720	720
160	3	906	1406	emb|Z26850|SPAT	S.pneumoniae (M222) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	95	501	501
160	4	1373	1942	emb|Z26850|SPAT	S.pneumoniae (M222) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	87	306	570
161	1	1	984	emb|X77249|SPR6	S.pneumoniae (R6) ciaR/ciaH genes	99	984	984
161	7	6910	7497	emb|X83917|SPGY	S.pneumoniae orf1gyrB and gyrB gene encoding DNA gyrase B subunit	99	437	588
161	8	7443	9386	emb|X83917|SPGY	S.pneumoniae orf1gyrB and gyrB gene encoding DNA gyrase B subunit	98	1912	1944
163	1	2	2155	gb|L20559|	Streptococcus pneumoniae Exp5 gene, partial cds	98	327	2154

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
165	1	32	1618	gb|J01796|	S.pneumoniae malX and malH genes encoding membrane protein and amylomaltase, complete cds, and malP gene encoding phosphorylase	99	1587	1587
165	2	1608	3902	gb|J01796|	S.pneumoniae malX and malH genes encoding membrane protein and amylomaltase, complete cds, and malP gene encoding phosphorylase	100	280	2295
166	1	378	4	emb|Y11463|SPDN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	100	375	375
166	2	1507	320	emb|Y11463|SPDN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	99	1188	1188
166	3	3240	1432	emb|Y11463|SPDN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	99	563	1809
167	1	1077	328	emb|Z71552|SPAD	Streptococcus pneumoniae adcCBA operon	94	155	750
167	2	1844	999	emb|Z71552|SPAD	Streptococcus pneumoniae adcCBA operon	98	405	846
167	3	2714	1842	emb|Z71552|SPAD	Streptococcus pneumoniae adcCBA operon	97	604	873
167	4	3399	2641	emb|Z71552|SPAD	Streptococcus pneumoniae adcCBA operon	99	703	759
168	1	1	2259	gb|L20558|	Streptococcus pneumoniae Exp4 gene, partial cds	99	282	2259
170	10	7338	7685	emb|Z77726|SPIS	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	95	315	348
172	6	2462	4981	gb|U47625|	Streptococcus pneumoniae formate acetyltransferase (exp72) gene, partial cds	97	365	2520
175	1	373	20	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	89	353	354
175	4	1843	3621	emb|Z47210|SPDE	S.pneumoniae dexB, cap3A, cap3B and cap3C genes and orfs	95	89	1779
176	5	3984	2980	emb|Z67739|SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	100	573	1005
178	1	3	425	emb|Z67739|SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	95	423	423
179	1	426	70	emb|Z83335|SPZ8	S.pneumoniae dexB, cap1[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	99	338	357
180	3	3084	1855	emb|X95718|SPGY	S.pneumoniae gyrA gene	99	381	1230
186	1	714	4	emb|Z79691|SOOR	S.pneumoniae yorf[A,B,C,D,E], ftsL, pbpX and regR genes	98	59	711
186	2	2254	608	emb|Z79691|SOOR	S.pneumoniae yorf[A,B,C,D,E], ftsL, pbpX and regR genes	98	315	1647
186	3	707	880	emb|Z79691|SOOR	S.pneumoniae yorf[A,B,C,D,E], ftsL, pbpX and regR genes	98	174	174
189	1	2	259	gb|U72720|	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	99	258	258
189	2	600	385	gb|U72720|	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	98	204	216

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
189	3	1018	851	gb|U72720|	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	99	168	168
189	4	1012	2154	gb|U72720|	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	99	1062	1143
191	9	7829	7524	emb|X63602|SPBO	S.pneumoniae mmsA-Box	95	234	306
194	1	1	729	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	91	728	729
199	2	1117	881	emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	96	211	237
199	4	1499	1762	emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	89	248	264
199	5	1781	2284	emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	98	504	504
203	1	1977	337	gb|L20563|	Streptococcus pneumoniae Exp3 gene, partial cds	99	342	1641
204	1	1145	3	gb|L36131|	Streptococcus pneumoniae exp10 gene, complete cds, recA gene, 5' end	99	1143	1143
208	1	59	2296	gb|U89711|	Streptococcus pneumoniae pneumococcal surface protein A PspA (pspA) gene, complete cds	90	471	2238
213	3	2455	2123	emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	96	332	333
216	1	360	12	emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	99	338	357
216	3	2650	2327	gb|M28678|	S.pneumoniae promoter sequence DNA	98	86	324
222	1	417	4	emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	94	414	414
227	3	5266	4238	emb|AJ000336|SP	Streptococcus pneumoniae ldh gene	99	1029	1029
239	1	1	804	gb|M31296|	S.pneumoniae recP gene, complete cds	95	484	804
247	3	1625	1807	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	94	178	183
249	3	921	1364	emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	94	443	444
253	1	362	3	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	360	360
253	5	1238	2050	emb|Z83335|SPZ8	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	95	420	813

TABLE 1 S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
253	6	2069	2572	emb|Z83335|SPZ8	S.pneumoniae dexB, capI[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	97	504	504
255	1	3	800	emb|Z82002|SPZ8	S.pneumoniae pcpB and pcpC genes	97	531	798
255	2	798	1841	emb|Z82002|SPZ8	S.pneumoniae pcpB and pcpC genes	97	672	1044
255	3	2493	1969	emb|Z67739|SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	92	435	525
257	2	985	770	emb|X17337|SPAM	Streptococcus pneumoniae ami locus conferring aminopterin resistance	96	117	216
257	3	1245	907	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	97	339	339
267	2	495	1208	gb|U16156|	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	95	84	714
267	3	1291	2277	gb|U16156|	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	97	755	987
267	4	2261	3601	gb|U16156|	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	98	1341	1341
267	5	3561	4136	gb|U16156|	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	99	576	576
267	6	4164	4949	gb|U16156|	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	99	748	786
267	7	5544	5140	gb|U16156|	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	100	186	405
268	4	1793	1990	emb|X63602|SPBO	S.pneumoniae mmsA-Box	89	194	198
271	1	562	104	gb|M29686|	S.pneumoniae mismatch repair (hexB) gene, complete cds	93	160	459
291	1	75	524	gb|U04047|	Streptococcus pneumoniae SSZ dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	96	450	450
291	2	1001	525	emb|Z83335|SPZ8	S pneumoniae dexB, capI[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	87	205	477
291	3	807	559	emb|Z83335|SPZ8	S.pneumoniae dexB, capI[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	90	170	249
291	4	1374	1099	gb|M36180|	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	85	264	276

TABLE 1 S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
293	1	3	1673	emb|Z67740|SPGY	S.pneumoniae gyrB gene and unknown orf	98	553	1671
296	1	1434	151	emb|Z47210|SPDE	S.pneumoniae dexB, cap1A, cap1B and cap1C genes and orfs	99	430	1284
317	1	157	510	emb|Z67739|SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	89	353	354
325	2	1237	485	emb|Z83335|SPZ8	S.pneumoniae dexB, cap1[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	91	299	753
326	1	1	462	emb|Z82001|SPZ8	S.pneumoniae pcpA gene and open reading frames	100	233	462
327	1	603	64	emb|Z83335|SPZ8	S.pneumoniae dexB, cap1[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	94	89	540
334	1	153	545	gb|U41735|	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	87	91	393
336	1	308	93	emb|Z26850|SPAT	S.pneumoniae (H222) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	97	102	216
360	1	1	519	emb|Z67739|SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	95	435	519
360	4	1598	1960	emb|Z83335|SPZ8	S.pneumoniae dexB, cap1[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	94	353	363
362	1	673	2	emb|Z83335|SPZ8	S.pneumoniae dexB, cap1[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose biosynthesis genes and aliA gene	95	63	672
362	2	1168	728	gb|U04047|	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	96	441	441
384	1	347	111	emb|X85787|SPCP	S.pneumoniae dexB, cps14A, cps14B, cps14C, cps14D, cps14E, cps14F, cps14G, cps14H, cps14I, cps14J, cps14K, cps14L, tasA genes	94	54	237

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
228	2	1760	1942	|pir|F60663|F606	|translation elongation factor Tu - Streptococcus oralis	100	100	183
319	1	2	205	|gi|984927	|neomycin phosphotransferase (cloning vector pBSL99)	100	100	204
260	1	2	1138	|pir|F60663|F606	|translation elongation factor Tu - Streptococcus oralis	99	98	1137
25	2	486	1394	|gi|1574495	|hypothetical (Haemophilus influenzae)	98	96	909
94	2	685	1002	|gi|310627	|phosphoenolpyruvate:sugar phosphotransferase system HPr (Streptococcus mutans)	98	93	318
312	1	190	2	|gi|347999	|ATP-dependent protease proteolytic subunit (Streptococcus salivarius)	98	95	189
329	1	1	807	|gi|924848	|inosine monophosphate dehydrogenase (Streptococcus pyogenes)	98	94	807
336	2	290	589	|gi|987050	|lacZ gene product (unidentified cloning vector)	98	98	300
181	9	5948	7366	|gi|153755	|phospho-beta-D-galactosidase (EC 3.2.1.85) (Lactococcus lactis cremoris)	97	94	1419
312	2	1044	361	|gi|347998	|uracil phosphoribosyltransferase (Streptococcus salivarius)	97	88	684
32	8	6575	7486	|sp|P37214|ERA_S	|GTP-BINDING PROTEIN ERA HOMOLOG	96	91	912
94	3	951	2741	|gi|153615	|phosphoenolpyruvate:sugar phosphotransferase system enzyme I (Streptococcus salivarius)	96	92	1791
127	1	1	168	|gi|581299	|initiation factor IF-1 (Lactococcus lactis)	96	89	168
128	14	10438	11154	|gi|1276873	|DeoD (Streptococcus thermophilus)	96	93	717
181	4	1362	1598	|gi|46606	|lacD polypeptide (AA 1-326) (Staphylococcus aureus)	96	80	237
218	1	1	834	|gi|1743856	|intrageneric coaggregation-relevant adhesin (Streptococcus gordonii)	96	93	834
319	2	115	441	|gi|208225	|heat-shock protein 82/neomcyn phosphotransferase fusion protein (hsp82-neo) (unidentified cloning vector)	96	96	327
54	12	8622	10967	|gnl|PID|d100972	|Pyruvate formate-lyase (Streptococcus mutans)	95	89	2346
181	2	606	1289	|gi|149396	|lacD (Lactococcus lactis)	95	89	684
46	3	3410	3045	|gi|1850606	|YlxM (Streptococcus mutans)	94	86	366
89	10	7972	7337	|gi|703442	|thymidine kinase (Streptococcus gordonii)	94	86	636
148	9	6431	7354	|gi|995767	|UDP-glucose pyrophosphorylase (Streptococcus pyogenes)	94	85	924
160	7	4430	5848	|gi|153573	|H+ ATPase (Enterococcus faecalis)	94	87	1419
2	3	4598	3513	|gi|153763	|plasmin receptor (Streptococcus pyogenes)	93	86	1086
12	8	7877	6204	|gi|1103865	|formyl-tetrahydrofolate synthetase (Streptococcus mutans)	93	84	1674

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
65	11	4734	5120	gi|40150	L14 protein (AA 1-122) [Bacillus subtilis]	93	87	387
68	1	53	1297	gi|47341	antitumor protein [Streptococcus pyogenes]	93	87	1245
80	1	3	299	gnl|PID|d101166	ribosomal protein S7 [Bacillus subtilis]	93	84	297
127	3	695	1093	gi|142462	ribosomal protein S11 [Bacillus subtilis]	93	86	399
160	5	1924	3462	gi|1773264	ATPase, alpha subunit [Streptococcus mutans]	93	85	1539
211	5	3757	3047	gi|535273	aminopeptidase C [Streptococcus thermophilus]	93	82	711
262	1	16	564	gi|149394	lacB [Lactococcus lactis]	93	90	549
366	1	197	3	gi|295259	tryptophan synthase beta subunit [Synechocystis sp.]	93	91	195
25	3	1392	1976	gi|1574496	hypothetical [Haemophilus influenzae]	92	80	585
36	21	20781	19927	gi|310632	hydrophobic membrane protein [Streptococcus gordonii]	92	86	855
181	3	1265	1534	gi|149396	lacD [Lactococcus lactis]	92	83	270
181	7	3662	4060	gi|149410	enzyme III [Lactococcus lactis]	92	83	399
32	4	5631	3937	gnl|PID|e294090	fibronectin-binding protein-like protein A [Streptococcus gordonii]	91	85	1695
46	2	3054	1462	gi|1850607	signal recognition particle Ffh [Streptococcus mutans]	91	84	1593
65	10	4442	4726	pir|S17865|S178	ribosomal protein S17 - Bacillus stearothermophilus	91	80	285
77	2	260	1900	gi|287871	groEL gene product [Lactococcus lactis]	91	82	1641
84	1	2	2056	gi|871784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	91	79	2055
99	8	10750	9272	gi|153740	sucrose phosphorylase [Streptococcus mutans]	91	84	1479
99	9	11947	11072	gi|153739	membrane protein [Streptococcus mutans]	91	78	876
127	5	2065	2469	pir|S07223|R5BS	ribosomal protein L17 - Bacillus stearothermophilus	91	78	405
132	6	9539	9390	gi|143065	hubst [Bacillus stearothermophilus]	91	89	150
137	8	4765	6153	gnl|PID|d100347	Na+ -ATPase beta subunit [Enterococcus hirae]	91	79	1389
151	7	11119	9734	gi|1815634	glutamine synthetase type 1 [Streptococcus agalactiae]	91	82	1386
201	2	1798	278	gi|2208998	dextran glucosidase DexS [Streptococcus suis]	91	79	1521
222	2	673	1839	gi|153741	ATP-binding protein [Streptococcus mutans]	91	85	1167
293	5	4113	4400	gi|1196921	unknown protein [Insertion sequence IS861]	91	71	288
32	7	6166	6570	pir|A36933|A369	diacylglycerol kinase homolog - Streptococcus mutans	90	77	405

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
33	2	841	527	gi|1196921	unknown protein (Insertion sequence IS861)	90	70	315
48	27	20908	19757	gnl|PID|e274705	lactate oxidase (Streptococcus lniae)	90	80	1152
55	21	19777	18515	gnl|PID|e221213	ClpX protein (Bacillus subtilis)	90	75	1263
56	2	717	977	gi|1710133	flagellar filament cap (Borrelia burgdorferi)	90	50	261
65	1	1	606	gi|1165303	L3 (Bacillus subtilis)	90	75	606
114	1	2	988	gi|153562	aspartate beta-semialdehyde dehydrogenase (EC 1.2.1.11) (Streptococcus mutans)	90	80	987
120	1	1345	827	gi|407880	ORF1 (Streptococcus equisimilis)	90	75	519
159	12	7690	8298	gi|143012	GMP synthetase (Bacillus subtilis)	90	84	609
166	4	4076	3282	gi|1661179	high affinity branched chain amino acid transport protein (Streptococcus mutans)	90	78	795
183	1	28	1395	gi|308858	ATP:pyruvate 2-O-phosphotransferase (Lactococcus lactis)	90	76	1368
191	3	2891	1662	gi|149521	tryptophan synthase beta subunit (Lactococcus lactis)	90	78	1230
198	2	1551	436	gi|2323342	(AF014460) CcpA (Streptococcus mutans)	90	76	1116
305	1	37	783	gi|1573551	asparagine synthetase A (asnA) (Haemophilus influenzae)	90	80	747
8	3	2285	3343	gi|149434	putative (Lactococcus lactis)	89	78	1059
46	8	7577	7362	pir|A45434|A454	ribosomal protein L19 - Bacillus stearothermophilus	89	76	216
49	9	8361	10342	gi|153792	recP peptide (Streptococcus pneumoniae)	89	83	1980
51	14	18410	19447	gi|308857	ATP:D-fructose 6-phosphate 1-phosphotransferase (Lactococcus lactis)	89	81	1038
57	11	9686	10669	gnl|PID|d100932	H2O-forming NADH Oxidase (Streptococcus mutans)	89	77	984
65	5	2418	2786	gi|1165307	S19 (Bacillus subtilis)	89	81	369
65	8	3806	4225	sp|P14577|RL16_	50S RIBOSOMAL PROTEIN L16.	89	82	420
65	18	8219	8719	gi|143417	ribosomal protein S5 (Bacillus stearothermophilus)	89	76	501
73	9	6337	5315	gi|532204	prs (Listeria monocytogenes)	89	70	1023
76	3	3360	1465	gnl|PID|e200671	lepA gene product (Bacillus subtilis)	89	76	1896
99	10	12818	11919	gi|153738	membrane protein (Streptococcus mutans)	89	73	900
120	2	1552	1300	gi|407881	stringent response-like protein (Streptococcus equisimilis)	89	79	2253
122	5	4512	2791	gnl|PID|e280490	unknown (Streptococcus pneumoniae)	89	81	1722

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
176	1	669	4	|gi|47394	|5-oxoprolyl-peptidase (Streptococcus pyogenes)	89	78	666
177	6	3050	3934	|gi|912423	|putative (Lactococcus lactis)	89	71	885
181	8	4033	5751	|gi|149411	|enzyme III (Lactococcus lactis)	89	80	1719
211	4	3149	2793	|gi|535273	|aminopeptidase C (Streptococcus thermophilus)	89	83	357
361	1	431	838	|gi|1196922	|unknown protein (Insertion sequence IS861)	89	70	408
34	17	11839	10535	|sp|P30053|SYH_S	|HISTIDYL-TRNA SYNTHETASE (EC 6.1.1.21) (HISTIDINE--TRNA LIGASE) (HISRS).	88	78	1305
38	3	1646	2623	|gi|2058544	|putative ABC transporter subunit ComYA (Streptococcus gordonii)	88	78	978
54	1	3	227	|gnl|PID|d101320	|YqgU (Bacillus subtilis)	88	66	225
57	2	611	1468	|gnl|PID|e134943	|putative reductase 1 (Saccharomyces cerevisiae)	88	75	858
65	13	5497	6069	|pir|A29102|R5BS	|ribosomal protein L5 - Bacillus stearothermophilus	88	75	573
65	20	9030	9500	|gi|2078381	|ribosomal protein L15 (Staphylococcus aureus)	88	83	471
78	3	3636	1108	|gnl|PID|d100781	|lysyl-aminopeptidase (Lactococcus lactis)	88	80	2529
106	12	12965	12054	|gi|2407215	|(AF017421) putative heat shock protein HtpX (Streptococcus gordonii)	88	72	912
107	2	219	962	|gnl|PID|e139862	|putative acylneuraminate lyase (Clostridium tertium)	88	75	744
111	8	14073	10420	|gi|402363	|RNA polymerase beta-subunit (Bacillus subtilis)	88	74	3654
126	9	13096	12062	|gnl|PID|e111468	|unknown (Bacillus subtilis)	88	74	1035
140	17	19143	18874	|gi|1573659	|H. influenzae predicted coding region HI0659 (Haemophilus influenzae)	88	61	270
144	1	394	555	|gnl|PID|e274705	|lactate oxidase (Streptococcus lniae)	88	75	162
148	4	2723	3493	|gi|1591672	|phosphate transport system ATP-binding protein (Methanococcus jannaschii)	88	68	771
160	8	5853	6278	|gi|1773267	|ATPase, epsilon subunit (Streptococcus mutans)	88	65	426
177	4	1770	2885	|gi|149426	|putative (Lactococcus lactis)	88	72	1116
211	6	4140	3613	|gi|535273	|aminopeptidase C (Streptococcus thermophilus)	88	74	528
231	4	580	957	|gi|40186	|homologous to E coli ribosomal protein L27 (Bacillus subtilis)	88	78	378
260	5	2387	2998	|gi|1196922	|unknown protein (Insertion sequence IS861)	88	69	612
291	6	2017	3375	|gnl|PID|d100571	|adenylosuccinate synthetase (Bacillus subtilis)	88	75	1359
319	4	658	317	|gi|603578	|serine/threonine kinase (Phytophthora capsici)	88	88	342
40	5	4353	4514	|gi|153672	|lactose repressor (Streptococcus mutans)	87	56	162

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
49	10	10660	10929	gi|1196921	unknown protein (Insertion sequence IS861)	87	72	270
65	7	3140	3808	gi|1165309	S3 (Bacillus subtilis)	87	73	669
65	15	6623	7039	gi|1044978	ribosomal protein S8 (Bacillus subtilis)	87	73	417
75	8	5411	6625	gi|1877422	galactokinase (Streptococcus mutans)	87	78	1215
80	2	703	2805	gnl|PID|d101166	elongation factor G (Bacillus subtilis)	87	76	2103
82	1	541	248	gi|1196921	unknown protein (Insertion sequence IS861)	87	69	294
140	23	25033	23897	gnl|PID|e254999	phenylalany-tRNA synthetase beta subunit (Bacillus subtilis)	87	74	1137
214	14	10441	8516	gi|2281305	glucose inhibited division protein homolog GidA (Lactococcus lactis cremoris)	87	75	1926
220	2	2742	874	gnl|PID|e324358	product highly similar to elongation factor EF-G (Bacillus subtilis)	87	73	1869
260	4	2096	2389	gi|1196921	unknown protein (Insertion sequence IS861)	87	72	294
323	1	27	650	gi|897795	30S ribosomal protein (Pediococcus acidilactici)	87	73	624
357	1	154	570	gi|1044978	ribosomal protein S8 (Bacillus subtilis)	87	73	417
49	11	10927	11445	gi|1196922	unknown protein (Insertion sequence IS861)	86	63	519
59	12	7461	9224	gi|951051	relaxase (Streptococcus pneumoniae)	86	68	1764
65	4	1553	2401	pir|A02759|R5BS	ribosomal protein L2 - Bacillus stearothermophilus	86	77	849
65	23	10957	11610	gi|44074	adenylate kinase (Lactococcus lactis)	86	76	654
82	4	4374	4856	gi|153745	mannitol-specific enzyme III (Streptococcus mutans)	86	72	483
102	4	4270	4986	gnl|PID|e264705	OMP decarboxylase (Lactococcus lactis)	86	76	717
106	6	7824	6880	gnl|PID|e337598	aspartate transcarbamylase (Lactobacillus leichmannii)	86	68	945
107	1	1	273	gnl|PID|e339862	putative acylneuraminate lyase (Clostridium tertium)	86	71	273
111	7	10432	6710	gnl|PID|e228283	DNA-dependent RNA polymerase (Streptococcus pyogenes)	86	80	3723
131	9	5704	4892	gi|1661193	polipoprotein diacylglycerol transferase (Streptococcus mutans)	86	71	813
134	7	6430	7980	gi|2388637	glycerol kinase (Enterococcus faecalis)	86	73	1551
146	11	7473	6583	gi|1591731	melvalonate kinase (Methanococcus jannaschii)	86	72	891
153	2	595	2010	gi|2160707	dipeptidase (Lactococcus lactis)	86	78	1416
154	1	2	1435	gi|1857246	6-phosphogluconate dehydrogenase (Lactococcus lactis)	86	74	1434

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
161	5	5025	6284	gi|47529	Unknown (Streptococcus salivarius)	86	66	1260
184	1	2	1483	gi|642667	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (Streptococcus mutans)	86	73	1482
210	8	3659	6571	gi|153661	translational initiation factor IF2 (Enterococcus faecium)	86	76	2913
250	1	2	187	gi|1573551	asparagine synthetase A (asnA) (Haemophilus influenzae)	86	68	186
36	4	2644	3909	gi|2149909	cell division protein (Enterococcus faecalis)	85	73	1266
38	4	2475	3587	gi|2058545	putative ABC transporter subunit ComYB (Streptococcus gordonii)	85	72	1113
38	5	3577	3915	gi|2058546	ComYC (Streptococcus gordonii)	85	80	339
57	5	2797	3789	gnl|PID|d101316	YqfJ (Bacillus subtilis)	85	72	993
82	5	4915	6054	gi|153746	mannitol-phosphate dehydrogenase (Streptococcus mutans)	85	68	1140
83	15	14690	15793	gi|143371	phosphoribosyl aminoimidazole synthetase (PUR-M) (Bacillus subtilis)	85	69	1104
87	2	1417	2388	gi|1184967	ScrR (Streptococcus mutans)	85	69	972
108	3	2666	3154	gi|153566	ORF (19K protein) (Enterococcus faecalis)	85	67	489
127	2	312	692	gi|1044989	ribosomal protein S13 (Bacillus subtilis)	85	72	381
128	3	1534	2409	gi|1685110	tetrahydrofolate dehydrogenase/cyclohydrolase (Streptococcus thermophilus)	85	71	876
137	7	2962	4767	gnl|PID|d100347	Na^+-ATPase alpha subunit (Enterococcus hirae)	85	74	1806
170	2	2622	709	gnl|PID|d102006	(AB001488) FUNCTION UNKNOWN, SIMILAR PRODUCT IN E.COLI, H. INFLUENZAE AND NEISSERIA MENINGITIDIS. (Bacillus subtilis)	85	70	1914
187	5	3760	4386	gi|727436	putative 20-kDa protein (Lactococcus lactis)	85	65	627
233	2	728	1873	gi|1163116	ORF-5 (Streptococcus pneumoniae)	85	67	1146
234	3	962	1255	gi|2293155	(AF008220) YtiA (Bacillus subtilis)	85	61	294
240	1	309	1931	gi|143597	CTP synthetase (Bacillus subtilis)	85	70	1623
6	1	199	1521	gi|508979	GTP-binding protein (Bacillus subtilis)	84	72	1323
10	4	4375	3443	gnl|PID|e339862	putative acylneuraminate lyase (Clostridium tertium)	84	70	933
14	1	63	2093	gi|520753	DNA topoisomerase I (Bacillus subtilis)	84	69	2031
19	4	1793	2593	gi|2352484	(AF005098) RNAseH II (Lactococcus lactis)	84	68	801
20	17	17720	19687	gnl|PID|d100584	cell division protein (Bacillus subtilis)	84	71	1968
22	28	21723	20884	gi|299163	alanine dehydrogenase (Bacillus subtilis)	84	68	840

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
30	10	7730	6792	|gnl|PID|d100296	|fructokinase (Streptococcus mutans)	84	75	939
33	9	5650	5300	|gi|147194	|phnA protein (Escherichia coli)	84	71	351
36	22	21551	20772	|gi|310631	|ATP binding protein (Streptococcus gordonii)	84	72	780
48	4	2837	2505	|gi|882609	|6-phospho-beta-glucosidase (Escherichia coli)	84	69	333
58	1	41	1516	|gi|450849	|amylase (Streptococcus bovis)	84	73	1476
59	10	6715	7116	|gi|951053	|ORF10, putative (Streptococcus pneumoniae)	84	74	402
62	1	21	644	|gi|806487	|ORF211; putative (Lactococcus lactis)	84	66	624
65	17	7779	8207	|gi|1044980	|ribosomal protein L18 (Bacillus subtilis)	84	73	429
65	21	9507	10397	|gi|44073	|SecY protein (Lactococcus lactis)	84	68	891
106	4	5474	2262	|gnl|PID|e199387	|carbamoyl-phosphate synthase (Lactobacillus plantarum)	84	73	3213
159	1	147	4	|gi|806487	|ORF211; putative (Lactococcus lactis)	84	63	144
163	4	4690	5910	|gi|2293164	|(AF008220) SAM synthase (Bacillus subtilis)	84	69	1221
192	1	46	1308	|gi|495046	|tripeptidase (Lactococcus lactis)	84	73	1263
348	1	671	6	|gi|1787753	|(AE000245) f346; 79 pct identical to 336 amino acids of ADH1_ZYMMO SW: P20368 but has 10 additional N-ter residues (Escherichia coli)	84	71	666
3	4	1572	3575	|gi|143766	|(thrSv) (EC 6.1.1.3) (Bacillus subtilis)	83	65	2004
9	6	3893	3417	|gnl|PID|d100576	|single strand DNA binding protein (Bacillus subtilis)	83	68	477
17	15	7426	8457	|gi|520738	|comA protein (Streptococcus pneumoniae)	83	66	1032
20	12	13860	14144	|gnl|PID|d100583	|unknown (Bacillus subtilis)	83	61	285
23	4	3358	2606	|gi|1788294	|(AE000290) o238; This 238 aa orf is 40 pct identical (5 gaps) to 231 residues of an approx. 248 aa protein YEBC_ECOLI SW: P24237 (Escherichia coli)	83	74	753
28	6	3304	3005	|gi|1573659	|H. influenzae predicted coding region HI0659 (Haemophilus influenzae)	83	57	300
35	7	5108	3867	|gi|311707	|hypothetical nucleotide binding protein (Acholeplasma laidlawii)	83	63	1242
55	19	17932	17528	|gi|537085	|ORF_f141 (Escherichia coli)	83	59	405
55	20	18539	17919	|gi|496558	|orfX (Bacillus subtilis)	83	69	621
65	6	2795	3142	|gi|1165308	|L22 (Bacillus subtilis)	83	64	348
68	6	6877	6683	|gi|1213494	|immunoglobulin A1 protease (Streptococcus pneumoniae)	83	54	195

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
87	15	15112	14771	gnl|PID|e323522	putative rpoZ protein (Bacillus subtilis)	83	54	342
96	12	8963	9631	gi|47394	5-oxoprolyl-peptidase (Streptococcus pyogenes)	83	73	669
98	1	3	263	gi|1183885	glutamine-binding subunit (Bacillus subtilis)	83	55	261
120	4	7170	5233	gi|310630	zinc metalloprotease (Streptococcus gordonii)	83	72	1938
127	7	2998	4347	gi|1500567	M. jannaschii predicted coding region MJ1665 (Methanococcus jannaschii)	83	72	1350
137	1	3	440	gi|472918	v-type Na-ATPase (Enterococcus hirae)	83	60	438
160	6	3466	4356	gi|1773265	ATPase, gamma subunit (Streptococcus mutans)	83	67	891
214	4	2278	2964	gi|663279	transposase (Streptococcus pneumoniae)	83	72	687
226	3	2367	2020	gi|142154	thioredoxin (Synechococcus PCC6301)	83	58	348
303	1	3	1049	gi|40046	phosphoglucose isomerase A (AA 1-449) (Bacillus stearothermophilus)	83	67	1047
303	2	1155	1931	gi|289282	glutamyl-tRNA synthetase (Bacillus subtilis)	83	67	777
6	17	15370	14318	gi|633147	ribose-phosphate pyrophosphokinase (Bacillus caldolyticus)	82	64	1053
7	1	299	96	gi|143648	ribosomal protein L28 (Bacillus subtilis)	82	69	204
9	3	1479	1090	gi|385178	unknown (Bacillus subtilis)	82	46	390
9	7	4213	3899	gnl|PID|d100576	ribosomal protein S6 (Bacillus subtilis)	82	60	315
12	6	4688	3942	gnl|PID|d100571	unknown (Bacillus subtilis)	82	68	747
22	17	13422	14837	gi|520754	putative (Bacillus subtilis)	82	69	1416
22	18	14897	15658	gnl|PID|d101929	uridine monophosphate kinase (Synechocystis sp.)	82	62	762
33	16	11471	10641	gnl|PID|d101190	ORF4 (Streptococcus mutans)	82	68	831
35	9	7400	6255	gi|1881543	UDP-N-acetylglucosamine-2-epimerase (Streptococcus pneumoniae)	82	68	1146
40	10	8003	7533	gi|1173519	riboflavin synthase beta subunit (Actinobacillus pleuropneumoniae)	82	68	471
48	32	23159	23437	gi|1930092	outer membrane protein (Campylobacter jejuni)	82	61	279
52	14	13833	14765	gi|142521	deoxyribodipyrimidine photolyase (Bacillus subtilis)	82	61	933
60	4	4737	1849	gnl|PID|d102221	(AB001810) uvrA (Deinococcus radiodurans)	82	66	2889
62	4	2131	1457	gi|2246749	(AF009622) thioredoxin reductase (Listeria monocytogenes)	82	63	675
71	11	16586	17518	gnl|PID|e322063	ss-1,4-galactosyltransferase (Streptococcus pneumoniae)	82	60	933
73	13	9222	7837	gnl|PID|d100586	unknown (Bacillus subtilis)	82	65	1386

WO 98/18931 PCT/US97/19588

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
74	1	1	3771	gnl|PID|d101199	alkaline amylopullulanase [Bacillus sp.]	82	68	3771
83	9	3696	3983	gnl|PID|e305362	unnamed protein product [Streptococcus thermophilus]	82	52	288
86	11	10776	9394	gi|683583	5-enolpyruvylshikimate-3-phosphate synthase [Lactococcus lactis]	82	67	1383
89	12	8295	9752	gi|40025	homologous to E.coli 50K [Bacillus subtilis]	82	66	1458
115	9	10347	8812	gnl|PID|d102090	(AB003927) phospho-beta-galactosidase 1 [Lactobacillus gasseri]	82	74	1536
118	1	1	1332	gnl|PID|d100579	seryl-tRNA synthetase [Bacillus subtilis]	82	71	1332
151	3	4657	6246	pir|S06097|S060	type I site-specific deoxyribonuclease (EC 3.1.21.3) CfrA chain S - Citrobacter freundii	82	66	1590
173	6	4183	3503	gi|2313836	(AE000584) conserved hypothetical protein [Helicobacter pylori]	82	68	681
177	12	5481	7442	gnl|PID|d101999	(AB001341) NcrB [Escherichia coli]	82	58	1962
193	2	178	576	pir|S08564|RJBS	ribosomal protein S9 - Bacillus stearothermophilus	82	70	399
245	2	258	845	gi|146402	EcoA type I restriction-modification enzyme S subunit [Escherichia coli]	82	68	588
9	5	3400	3146	gnl|PID|d100576	ribosomal protein S18 [Bacillus subtilis]	81	66	255
16	7	7484	8413	gi|1100074	tryptophanyl-tRNA synthetase [Clostridium longisporum]	81	70	930
20	11	10308	13820	gnl|PID|d100583	transcription-repair coupling factor [Bacillus subtilis]	81	63	3513
38	2	1232	1606	gi|2058543	putative DNA binding protein [Streptococcus gordonii]	81	63	375
45	2	3061	1751	gi|460259	enolase [Bacillus subtilis]	81	67	1311
46	1	2	1267	gi|431231	uracil permease [Bacillus caldolyticus]	81	61	1266
48	3	2453	1440	gnl|PID|d100453	Mannosephosphate Isomerase [Streptococcus mutans]	81	70	1014
54	2	1106	336	gi|154752	transport protein [Agrobacterium tumefaciens]	81	64	771
65	22	10306	10821	gi|44073	SecY protein [Lactococcus lactis]	81	66	516
89	4	3874	2603	gi|556886	serine hydroxymethyltransferase [Bacillus subtilis]	81	69	1272
99	16	19126	18929	gi|2313526	(AE000557) H. pylori predicted coding region HP0411 [Helicobacter pylori]	81	75	198
106	7	8373	7822	gnl|PID|e199384	pyrR [Lactobacillus plantarum]	81	61	552
108	6	5054	6877	gi|1469939	group B oligopeptidase PepB [Streptococcus agalactiae]	81	66	1824
113	15	15899	18281	pir|S09411|S094	spoIIIE protein - Bacillus subtilis	81	65	2385
128	5	3359	3634	gi|1685111	orf1091 [Streptococcus thermophilus]	81	69	276

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
151	1	830	3211	gi|304896	EcoE type I restriction-modification enzyme R subunit (Escherichia coli)	81	59	2382
159	11	6722	7837	gi|2239288	GMP synthetase (Bacillus subtilis)	81	69	1116
170	1	739	458	gnl|PID|d102006	(AB001488) FUNCTION UNKNOWN. (Bacillus subtilis)	81	55	282
191	2	1759	893	gi|149522	tryptophan synthase alpha subunit (Lactococcus lactis)	81	65	867
214	3	2290	1994	gi|157587	reverse transcriptase endonuclease (Drosophila virilis)	81	43	297
217	4	4415	4008	gi|466473	cellobiose phosphotransferase enzyme II' (Bacillus stearothermophilus)	81	59	408
262	2	569	868	gi|153675	tagatose 6-P kinase (Streptococcus mutans)	81	68	300
299	1	663	4	gnl|PID|e301154	StySKI methylase (Salmonella enterica)	81	60	660
366	2	376	83	gi|149521	tryptophan synthase beta subunit (Lactococcus lactis)	81	65	294
12	10	8766	9242	gi|1216490	DNA/pantothenate metabolism flavoprotein (Streptococcus mutans)	80	64	477
17	11	6050	5748	gnl|PID|e305362	unnamed protein product (Streptococcus thermophilus)	80	67	303
17	16	8455	9066	gi|703126	leucocin A translocator (Leuconostoc gelidum)	80	59	612
18	1	2440	1613	gi|1591672	phosphate transport system ATP-binding protein (Methanococcus jannaschii)	80	58	828
27	3	4248	1579	gi|452309	valyl-tRNA synthetase (Bacillus subtilis)	80	69	2670
28	7	3671	3288	gi|1573660	H. influenzae predicted coding region HI0660 (Haemophilus influenzae)	80	63	384
32	2	902	1933	gnl|PID|e264499	dihydroorotate dehydrogenase B (Lactococcus lactis)	80	66	1032
39	1	1	1266	gnl|PID|e234078	hom (Lactococcus lactis)	80	63	1266
52	5	4363	3593	gi|1183884	ATP-binding subunit (Bacillus subtilis)	80	57	771
54	5	4550	4744	gi|2198820	(AF004225) Cux/CDP(1B1); Cux/CDP homeoprotein (Mus musculus)	80	60	195
59	11	7109	7486	gi|951052	ORF9, putative (Streptococcus pneumoniae)	80	68	378
65	3	1230	1550	pir|A02815|R5BS	ribosomal protein L23 - Bacillus stearothermophilus	80	69	321
65	12	5174	5503	pir|A02819|R5BS	ribosomal protein L24 - Bacillus stearothermophilus	80	70	330
66	9	9884	10687	gi|2313836	(AE000584) conserved hypothetical protein (Helicobacter pylori)	80	66	804
82	2	648	2438	gi|622991	mannitol transport protein (Bacillus stearothermophilus)	80	65	1791
85	1	950	630	gi|528995	polyketide synthase (Bacillus subtilis)	80	46	321
89	8	6870	5779	gi|853776	peptide chain release factor 1 (Bacillus subtilis)	80	63	1092
93	12	8718	7438	gnl|PID|d101959	hypothetical protein (Synechocystis sp.)	80	60	1281

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
106	5	6854	5751	|gnl|PID|e199386	|glutaminase of carbamoyl-phosphate synthase (Lactobacillus plantarum)	80	65	1104
109	2	2160	1450	|gi|40056	|phoP gene product (Bacillus subtilis)	80	59	711
124	9	4246	3953	|gnl|PID|d102254	|30S ribosomal protein S16 (Bacillus subtilis)	80	65	294
128	8	5148	6428	|gi|2281308	|phosphopentomutase (Lactococcus lactis cremoris)	80	66	1281
137	19	12665	11376	|gi|159109	|NADP-dependent glutamate dehydrogenase (Giardia intestinalis)	80	68	1290
140	19	19699	19457	|gi|517210	|putative transposase (Streptococcus pyogenes)	80	70	243
158	2	2474	984	|gi|1877423	|galactose-1-P-uridyl transferase (Streptococcus mutans)	80	65	1491
171	10	7474	7728	|gi|397800	|cyclophilin C-associated protein (Mus musculus)	80	60	255
181	1	2	619	|gi|149395	|lacC (Lactococcus lactis)	80	66	618
313	1	27	539	|gi|143467	|ribosomal protein S4 (Bacillus subtilis)	80	70	513
329	2	1652	858	|gi|533080	|RecF protein (Streptococcus pyogenes)	80	63	795
371	1	2	958	|gi|442360	|ClpC adenosine triphosphatase (Bacillus subtilis)	80	58	957
8	7	4312	5580	|gi|149435	|putative (Lactococcus lactis)	79	64	1269
23	1	1175	135	|gi|1542975	|AbcB (Thermoanaerobacterium thermosulfurigenes)	79	61	1041
33	14	9244	8201	|gnl|PID|e253891	|UDP-glucose 4-epimerase (Bacillus subtilis)	79	62	1044
36	3	1242	2633	|gnl|PID|e324218	|ftsA (Enterococcus hirae)	79	58	1392
38	13	7155	8378	|gi|405134	|acetate kinase (Bacillus subtilis)	79	58	1224
55	7	9011	8229	|gi|1146234	|dihydrodipicolinate reductase (Bacillus subtilis)	79	56	783
65	19	8661	8915	|gi|2078380	|ribosomal protein L30 (Staphylococcus aureus)	79	68	255
69	4	3678	2128	|gnl|PID|e311452	|unknown (Bacillus subtilis)	79	64	1551
69	9	7881	7279	|gi|677850	|hypothetical protein (Staphylococcus aureus)	79	59	603
72	10	8491	9783	|gnl|PID|d101091	|hypothetical protein (Synechocystis sp.)	79	62	1293
80	3	2906	7300	|gi|143342	|polymerase III (Bacillus subtilis)	79	65	4395
82	14	13326	15689	|gnl|PID|e255093	|hypothetical protein (Bacillus subtilis)	79	65	2364
86	13	12233	11118	|gi|683582	|prephenate dehydrogenase (Lactococcus lactis)	79	58	1116
92	3	940	1734	|gi|537286	|triosephosphate isomerase (Lactococcus lactis)	79	65	795
98	6	4023	4742	|gnl|PID|d100262	|LivG protein (Salmonella typhimurium)	79	63	720

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
99	12	16315	14150	gi|153736	a-galactosidase (Streptococcus mutans)	79	64	2166
107	7	5684	6406	gi|460080	D-alanine:D-alanine ligase-related protein (Enterococcus faecalis)	79	58	723
113	9	6858	8303	gi|466882	ppsI; B1496_C2_189 (Mycobacterium leprae)	79	64	1446
151	10	13424	12213	gi|450686	3-phosphoglycerate kinase (Thermotoga maritima)	79	60	1212
162	2	1158	3017	gi|506700	CapD (Staphylococcus aureus)	79	67	1860
177	5	2876	3052	gi|912423	putative (Lactococcus lactis)	79	61	177
177	8	4198	4563	gi|149429	putative (Lactococcus lactis)	79	61	366
187	3	2728	2907	gnl|PID|d102002	(AB001488) FUNCTION UNKNOWN. (Bacillus subtilis)	79	53	180
189	7	3589	4350	gnl|PID|e183449	putative ATP-binding protein of ABC-type (Bacillus subtilis)	79	61	762
191	5	4249	3449	gi|149519	indoleglycerol phosphate synthase (Lactococcus lactis)	79	66	801
211	3	1805	2737	gi|147404	mannose permease subunit II-M-Man (Escherichia coli)	79	57	933
212	3	3863	3621	gnl|PID|e209004	glutaredoxin-like protein (Lactococcus lactis)	79	58	243
215	1	987	715	gi|2293242	(AF008220) arginine succinate synthase (Bacillus subtilis)	79	64	273
323	2	510	781	gi|897795	30S ribosomal protein (Pediococcus acidilactici)	79	67	252
380	1	694	2	gi|1184680	polynucleotide phosphorylase (Bacillus subtilis)	79	64	693
384	2	655	239	gi|143328	phoP protein (put.); putative (Bacillus subtilis)	79	59	417
6	3	2820	4091	gi|853767	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (Bacillus subtilis)	78	62	1272
8	1	50	1786	gi|149432	putative (Lactococcus lactis)	78	63	1737
9	1	351	124	gi|897793	y98 gene product (Pediococcus acidilactici)	78	59	228
15	8	7364	8314	gnl|PID|d100585	cysteine synthetase A (Bacillus subtilis)	78	63	951
20	10	9738	10310	gnl|PID|d100583	stage V sporulation (Bacillus subtilis)	78	58	573
20	16	17165	17713	gi|49105	hypoxanthine phosphoribosyltransferase (Lactococcus lactis)	78	59	549
22	22	17388	18416	gnl|PID|d101315	YqfE (Bacillus subtilis)	78	60	1029
22	27	20971	20612	gi|299163	alanine dehydrogenase (Bacillus subtilis)	78	59	360
34	8	7407	7105	gi|41015	aspartate-tRNA ligase (Escherichia coli)	78	55	303
35	8	6257	5196	gi|1657644	Cap8E (Staphylococcus aureus)	78	60	1062

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
40	11	9287	8001	gi|1173518	GTP cyclohydrase II/ 3,4-dihydroxy 2-butanone-4-phosphate synthase (Actinobacillus pleuropneumoniae)	78	58	1287
48	31	22422	23183	gi|2314330	(AE000623) glutamine ABC transporter, ATP-binding protein (glnQ) (Helicobacter pylori)	78	58	762
52	2	2101	1430	gi|1183887	integral membrane protein (Bacillus subtilis)	78	54	672
55	14	13605	12712	gnl|PID|d102026	(AB002150) YbbP (Bacillus subtilis)	78	58	894
55	17	16637	15612	gnl|PID|e313027	hypothetical protein (Bacillus subtilis)	78	51	1026
71	14	19756	19598	gi|179764	calcium channel alpha-1D subunit (Homo sapiens)	78	57	159
74	11	15031	14018	gi|1573279	Holliday junction DNA helicase (ruvB) (Haemophilus influenzae)	78	57	1014
75	9	6623	7972	gi|1877423	galactose-1-P-uridyl transferase (Streptococcus mutans)	78	62	1350
81	12	12125	13906	gi|1573607	L-fucose isomerase (fucI) (Haemophilus influenzae)	78	66	1782
82	3	2423	4417	gi|153744	ORF X; putative (Streptococcus mutans)	78	64	1995
83	18	16926	18500	gi|143373	phosphoribosyl aminoimidazole carboxy formyl formyltransferase/inosine monophosphate cyclohydrolase (PUR-H(J)) (Bacillus subtilis)	78	63	1575
83	20	20212	20775	gi|143364	phosphoribosyl aminoimidazole carboxylase I (PUR-E) (Bacillus subtilis)	78	64	564
92	2	165	878	gnl|PID|d101190	ORF2 (Streptococcus mutans)	78	62	714
98	8	5863	6909	gi|2331287	(AF013188) release factor 2 (Bacillus subtilis)	78	63	1047
113	3	1071	2741	gi|580914	dnaZX (Bacillus subtilis)	78	64	1671
127	4	1133	2071	gi|142463	RNA polymerase alpha-core-subunit (Bacillus subtilis)	78	59	939
132	1	2782	497	gi|1561763	pullulanase (Bacteroides thetaiotaomicron)	78	58	2286
135	4	2698	3537	gi|1788036	(AE000269) NH3-dependent NAD synthetase (Escherichia coli)	78	66	840
140	24	26853	25423	gi|1100077	phospho-beta-glucosidase (Clostridium longisporum)	78	64	1431
150	5	4690	4514	gi|149464	amino peptidase (Lactococcus lactis)	78	42	177
152	1	1	795	gi|639915	NADH dehydrogenase subunit (Thunbergia alata)	78	43	795
162	4	4997	4110	gnl|PID|e323528	putative YhaP protein (Bacillus subtilis)	78	64	888
181	10	8651	7947	gi|149402	lactose repressor (lacR; alt.) (Lactococcus lactis)	78	48	705
200	4	3627	4958	gnl|PID|d100172	invertase (Zymomonas mobilis)	78	61	1332
203	3	3230	3015	gi|1174237	CycK (Pseudomonas fluorescens)	78	57	216

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
210	9	6789	7172	|gi|580902	|ORF6 gene product (Bacillus subtilis)	78	42	384
214	6	3810	2797	|gnl|PID|d102049	|P. haemolytica o-sialoglycoprotein endopeptidase; P36175 (660) transmembrane (Bacillus subtilis)	78	60	1014
214	13	6322	8163	|gi|1377831	|unknown (Bacillus subtilis)	78	62	1842
217	1	9	2717	|gi|488430	|alcohol dehydrogenase 2 (Entamoeba histolytica)	78	64	2709
222	3	2316	3098	|gi|1573047	|spore germination and vegetative growth protein (gerC2) (Haemophilus influenzae)	78	65	783
268	1	742	8	|gi|517210	|putative transposase (Streptococcus pyogenes)	78	65	735
276	1	223	753	|gnl|PID|d100306	|ribosomal protein L1 (Bacillus subtilis)	78	65	531
312	3	1567	1079	|gi|289261	|comE ORF2 (Bacillus subtilis)	78	54	489
339	1	117	794	|gi|1916729	|CadD (Staphylococcus aureus)	78	53	678
342	2	762	265	|gi|1842439	|phosphatidylglycerophosphate synthase (Bacillus subtilis)	78	59	498
383	1	737	3	|gi|1184680	|polynucleotide phosphorylase (Bacillus subtilis)	78	64	735
7	15	11923	11018	|gi|1399855	|carboxyltransferase beta subunit (Synechococcus PCC7942)	77	63	906
8	2	1698	2255	|gi|149433	|putative (Lactococcus lactis)	77	59	558
17	14	6948	7550	|gi|520738	|comA protein (Streptococcus pneumoniae)	77	60	603
30	12	9761	8967	|gi|1000451	|TreP (Bacillus subtilis)	77	43	795
36	14	11421	12131	|gi|1573766	|phosphoglyceromutase (gpmA) (Haemophilus influenzae)	77	64	711
55	3	3836	4096	|gi|1708640	|YeaB (Bacillus subtilis)	77	55	261
61	8	8377	8054	|gi|1890649	|multidrug resistance protein LmrA (Lactococcus lactis)	77	51	324
65	2	607	1254	|gi|40103	|ribosomal protein L4 (Bacillus stearothermophilus)	77	63	648
68	8	7509	7240	|gi|47551	|MRP (Streptococcus suis)	77	68	270
69	1	1083	118	|gnl|PID|e311493	|unknown (Bacillus subtilis)	77	57	966
77	5	4583	4026	|gnl|PID|e281578	|hypothetical 12.2 kd protein (Bacillus subtilis)	77	60	558
83	14	13104	14552	|gi|1590947	|amidophosphoribosyltransferase (Methanococcus jannaschii)	77	56	1449
94	4	3006	5444	|gnl|PID|e329895	|(AJ000496) cyclic nucleotide-gated channel beta subunit (Rattus norvegicus)	77	66	2439
96	11	8518	8880	|gi|551879	|ORF 1 (Lactococcus lactis)	77	62	363
99	11	14082	12799	|gi|153737	|sugar-binding protein (Streptococcus mutans)	77	61	1284

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
106	2	361	1176	gi|148921	LicD protein [Haemophilus influenzae]	77	51	816
108	4	3152	4030	gi|1574730	tellurite resistance protein (tehB) [Haemophilus influenzae]	77	58	879
118	4	3520	3131	gi|1573900	D-alanine permease (dagA) [Haemophilus influenzae]	77	57	390
124	4	1796	1071	gi|1573162	tRNA (guanine-N1)-methyltransferase (trmD) [Haemophilus influenzae]	77	58	726
126	4	5909	4614	gnl|PID|d101163	Srb [Bacillus subtilis]	77	62	1296
128	2	630	1373	gnl|PID|d101328	YqiZ [Bacillus subtilis]	77	58	744
130	1	1	1287	gnl|PID|e325013	hypothetical protein [Bacillus subtilis]	77	61	1287
139	5	4388	3639	gi|2293302	(AF008220) YtqA [Bacillus subtilis]	77	59	750
140	11	10931	9582	gi|289284	cysteinyl-tRNA synthetase [Bacillus subtilis]	77	64	1350
140	18	19451	19263	gi|517210	putative transposase [Streptococcus pyogenes]	77	66	189
141	2	976	1683	gnl|PID|e157887	URF5 (aa 1-573) [Drosophila yakuba]	77	50	708
141	4	2735	5293	gi|556258	secA [Listeria monocytogenes]	77	59	2559
144	2	671	2173	gnl|PID|d100585	lysyl-tRNA thynthetase [Bacillus subtilis]	77	61	1503
163	5	6412	7398	gi|511015	dihydroorotate dehydrogenase A [Lactococcus lactis]	77	62	987
164	10	7841	7074	gnl|PID|d100964	homologue of iron dicitrate transport ATP-binding protein FecE of E. coli [Bacillus subtilis]	77	52	768
191	8	7257	5791	gi|149516	anthranilate synthase alpha subunit [Lactococcus lactis]	77	57	1467
198	8	5177	5177	gi|1573856	hypothetical [Haemophilus influenzae]	77	66	201
213	1	202	462	gi|1743860	Brca2 [Mus musculus]	77	50	261
250	2	231	509	gnl|PID|e334776	YlbH protein [Bacillus subtilis]	77	60	279
289	3	1737	1276	gnl|PID|d100947	Ribosomal Protein L10 [Bacillus subtilis]	77	62	462
292	2	1399	668	gi|143004	transfer RNA-Gln synthetase [Bacillus stearothermophilus]	77	58	732
7	3	2734	1166	gnl|PID|d101824	peptide-chain-release factor 3 [Synechocystis sp.]	76	53	1569
7	23	18474	18235	gi|455157	acyl carrier protein [Cryptomonas phi]	76	57	240
9	8	5706	4342	gi|1146247	asparaginyl-tRNA synthetase [Bacillus subtilis]	76	61	1365
10	5	4531	4185	gnl|PID|e314495	hypothetical protein [Clostridium perfringens]	76	53	147
18	2	1615	842	gi|1591672	phosphate transport system ATP-binding protein [Methanococcus jannaschii]	76	56	774

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
22	37	27796	28173	gnl|PID|e13389	translation initiation factor IF3 (AA 1-172) [Bacillus stearothermophilus]	76	64	378
35	6	3869	2682	gi|1771346	CapSG [Staphylococcus aureus]	76	61	1188
48	28	21113	21787	gi|2314328	(AE000623) glutamine ABC transporter, permease protein (glnP) [Helicobacter pylori]	76	52	675
52	12	12881	13786	gi|142521	deoxyribodipyrimidine photolyase [Bacillus subtilis]	76	58	906
55	10	11521	10571	gnl|PID|e283110	femD [Staphylococcus aureus]	76	61	951
57	8	7824	6559	gi|290561	o188 [Escherichia coli]	76	47	1266
62	5	2406	2095	gnl|PID|e313024	hypothetical protein [Bacillus subtilis]	76	59	312
65	9	4223	4441	gi|40148	L29 protein (AA 1-66) [Bacillus subtilis]	76	58	219
68	2	1328	2371	gnl|PID|e284233	anabolic ornithine carbamoyltransferase [Lactobacillus plantarum]	76	61	1044
69	8	7297	6005	gnl|PID|d101420	Pyrimidine nucleoside phosphorylase [Bacillus stearothermophilus]	76	61	1293
73	12	7839	7267	gnl|PID|e243629	unknown [Mycobacterium tuberculosis]	76	53	573
74	5	8433	7039	gnl|PID|d102048	C. thermocellum beta-glucosidase; P26208 (985) [Bacillus subtilis]	76	60	1395
80	5	7643	7916	gi|2314010	(AE000599) conserved hypothetical protein [Helicobacter pylori]	76	61	294
82	15	16019	16996	gi|1573900	D-alanine permease (dagA) [Haemophilus influenzae]	76	56	978
83	19	18616	19884	gi|143374	phosphoribosyl glycinamide synthetase (PUR-D; gtg start codon) [Bacillus subtilis]	76	60	1269
86	14	13409	12231	gi|143806	AroF [Bacillus subtilis]	76	58	1179
87	1	3	1442	gi|153804	sucrose-6-phosphate hydrolase [Streptococcus mutans]	76	59	1440
87	16	15754	15110	gnl|PID|e323500	putative Gmk protein [Bacillus subtilis]	76	56	645
93	4	1769	1539	gi|1574820	1,4-alpha-glucan branching enzyme (glgB) [Haemophilus influenzae]	76	46	231
94	1	51	365	gi|144313	6.0 kd ORF [Plasmid ColE1]	76	73	315
116	2	2151	1678	gi|153841	pneumococcal surface protein A [Streptococcus pneumoniae]	76	59	474
123	6	3442	5895	gi|1314297	ClpC ATPase [Listeria monocytogenes]	76	59	2454
126	2	2156	2932	gnl|PID|d101328	Yqi2 [Bacillus subtilis]	76	61	777
128	10	6973	7797	gi|944944	purine nucleoside phosphorylase [Bacillus subtilis]	76	60	825
131	11	6186	5812	gi|1674310	(AE000058) Mycoplasma pneumoniae, MG085 homolog, from M. genitalium [Mycoplasma pneumoniae]	76	47	375

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
139	4	3641	3192	gi|2293302	(AF008220) YtqA (Bacillus subtilis)	76	53	450
140	14	14872	12536	gi|1184680	polynucleotide phosphorylase (Bacillus subtilis)	76	62	2337
143	2	2583	3905	gi|143795	transfer RNA-Tyr synthetase (Bacillus subtilis)	76	61	1323
170	6	5095	6114	gnl|PID|d100959	ycgQ (Bacillus subtilis)	76	44	1020
180	2	1927	557	gi|40019	ORF 821 (aa 1-821) (Bacillus subtilis)	76	53	1371
191	7	5815	5228	gi|551880	anthranilate synthase beta subunit (Lactococcus lactis)	76	61	588
195	3	3829	2444	gi|2149905	D-glutamic acid adding enzyme (Enterococcus faecalis)	76	60	1386
200	3	1914	3629	gi|431272	lysis protein (Bacillus subtilis)	76	58	1716
201	1	431	207	gi|2208998	dextran glucosidase DexS (Streptococcus suis)	76	57	225
214	2	1283	2380	gi|663278	transposase (Streptococcus pneumoniae)	76	55	1098
225	3	2338	3411	gi|1552775	ATP-binding protein (Escherichia coli)	76	56	1074
233	1	2	724	gi|1163115	neuraminidase B (Streptococcus pneumoniae)	76	60	723
347	1	523	38	gi|537033	ORF_f356 (Escherichia coli)	76	60	486
356	2	842	165	gi|2149905	D-glutamic acid adding enzyme (Enterococcus faecalis)	76	61	678
366	3	734	348	gi|149520	phosphoribosyl anthranilate isomerase (Lactococcus lactis)	76	69	387
5	8	12599	11484	gi|1574293	fimbrial transcription regulation repressor (pilB) (Haemophilus influenzae)	75	61	1116
6	13	12553	11894	gnl|PID|d102050	ydiH (Bacillus subtilis)	75	51	660
9	10	7282	6062	gi|142538	aspartate aminotransferase (Bacillus sp.)	75	55	1221
10	12	8080	7940	gi|149493	SCRFI methylase (Lactococcus lactis)	75	56	141
18	5	4266	3301	gnl|PID|d101319	YqgH (Bacillus subtilis)	75	52	966
22	4	1838	2728	gi|1373157	orf-X; hypothetical protein; Method: conceptual translation supplied by author (Bacillus subtilis)	75	62	891
30	11	9015	7828	gi|153801	enzyme scr-II (Streptococcus mutans)	75	64	1188
31	5	2362	2010	gi|2293211	(AF008220) putative thioredoxin (Bacillus subtilis)	75	53	333
32	9	7484	8359	gnl|PID|d100560	formamidopyrimidine-DNA glycosylase (Streptococcus mutans)	75	61	876
33	4	1735	1448	gi|413976	lpa-52r gene product (Bacillus subtilis)	75	53	288
33	10	6470	5769	gi|533105	unknown (Bacillus subtilis)	75	56	702

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
33	12	6878	7183	pir|A00205|FECL	ferredoxin (4Fe-4S) - Clostridium thermaceticum	75	56	306
36	1	181	2	gi|2088739	(AF003141) strong similarity to the FABP/P2/CRBP/CRABP family of transporters (Caenorhabditis elegans)	75	43	180
38	22	14510	15379	gi|1574058	hypothetical (Haemophilus influenzae)	75	56	870
48	33	23398	24066	gi|1930092	outer membrane protein (Campylobacter jejuni)	75	56	669
51	1	2	319	gi|43985	nifS-like gene (Lactobacillus delbrueckii)	75	55	318
51	10	8318	11683	gi|537192	CG Site No. 620; alternate gene names hs, hsp, hsr, rmx apparent frameshift in GenBank Accession Number X06545 (Escherichia coli)	75	50	3366
54	18	19566	20759	gi|666069	orf2 gene product (Lactobacillus leichmannii)	75	58	1194
57	9	8448	7822	gi|290561	o188 (Escherichia coli)	75	50	627
65	14	6072	6356	gi|606241	30S ribosomal subunit protein S14 (Escherichia coli)	75	64	285
70	4	3071	2472	gi|1256617	adenine phosphoribosyltransferase (Bacillus subtilis)	75	57	600
71	24	30399	29404	gi|1574390	C4-dicarboxylate transport protein (Haemophilus influenzae)	75	57	996
73	2	910	455	gnl|PID|e249656	YneT (Bacillus subtilis)	75	57	456
79	1	1810	491	gi|1146219	28.2% of identity to the Escherichia coli GTP-binding protein Era; putative (Bacillus subtilis)	75	59	1320
82	6	6360	6536	gi|1655715	BztD (Rhodobacter capsulatus)	75	55	177
83	6	1938	2975	gnl|PID|e323529	putative PlsX protein (Bacillus subtilis)	75	56	1038
93	11	7368	5317	gi|39989	methionyl-tRNA synthetase (Bacillus stearothermophilus)	75	58	2052
93	13	9409	8699	gi|1591493	glutamine transport ATP-binding protein Q (Methanococcus jannaschii)	75	54	711
95	1	1795	47	gnl|PID|e323510	YloV protein (Bacillus subtilis)	75	57	1749
103	2	362	1186	gnl|PID|e266928	unknown (Mycobacterium tuberculosis)	75	64	825
104	1	691	915	gi|460026	repressor protein (Streptococcus pneumoniae)	75	54	225
113	5	2951	3883	gnl|PID|d101119	ABC transporter subunit (Synechocystis sp.)	75	55	933
121	1	320	1390	gi|2145131	repressor of class I heat shock gene expression HrcA (Streptococcus mutans)	75	58	1071
127	6	2614	3000	gi|1500451	M. jannaschii predicted coding region MJ1558 (Methanococcus jannaschii)	75	44	387
137	18	10082	10687	gi|393116	P-glycoprotein 5 (Entamoeba histolytica)	75	52	606
149	11	8499	9338	gnl|PID|d100582	unknown (Bacillus subtilis)	75	55	840

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
151	6	9100	7673	|gi|40467	|HsdS polypeptide, part of CfrA family |Citrobacter freundii|	75	57	1428
158	1	986	3	|gnl|PID|e253891	|UDP-glucose 4-epimerase |Bacillus subtilis|	75	63	984
172	8	5653	6774	|gi|142978	|glycerol dehydrogenase |Bacillus stearothermophilus|	75	56	1122
172	9	7139	9730	|gnl|PID|e268456	|unknown |Mycobacterium tuberculosis|	75	58	2592
173	1	261	79	|gnl|PID|e236469	|C10C5.6 |Caenorhabditis elegans|	75	50	183
185	3	3066	2014	|gi|1574806	|spermidine/putrescine transport ATP-binding protein (potA) |Haemophilus influenzae|	75	56	1053
191	6	5235	4213	|gi|149518	|phosphoribosyl anthranilate transferase |Lactococcus lactis|	75	61	1023
226	2	1774	1181	|gi|2314588	|(AE000642) conserved hypothetical protein |Helicobacter pylori|	75	65	594
231	1	1	153	|gi|40173	|homolog of E.coli ribosomal protein L21 |Bacillus subtilis|	75	57	153
234	1	2	418	|gi|2293259	|(AF008220) YtqI |Bacillus subtilis|	75	59	417
279	1	552	151	|gi|1119198	|unknown protein |Bacillus subtilis|	75	50	402
291	7	3558	3827	|gi|40011	|ORF17 (AA 1-161) |Bacillus subtilis|	75	48	270
375	2	137	628	|gi|410137	|ORFX1 |Bacillus subtilis|	75	58	492
6	20	16721	17560	|gi|2293323	|(AF008220) YtdI |Bacillus subtilis|	74	53	840
7	6	4682	6052	|gi|1354211	|PET112-like protein |Bacillus subtilis|	74	60	1371
18	4	3141	2427	|gnl|PID|d101319	|YqgI |Bacillus subtilis|	74	54	915
21	6	5885	4800	|gi|1072381	|glutamyl-aminopeptidase |Lactococcus lactis|	74	59	1086
24	2	739	548	|gi|2314762	|(AE000655) ABC transporter, permease protein (yaeE) |Helicobacter pylori|	74	46	192
25	1	2	367	|gnl|PID|d100932	|H2O-forming NADH Oxidase |Streptococcus mutans|	74	63	366
38	18	11432	12964	|gi|537034	|ORF_o488 |Escherichia coli|	74	57	1533
48	10	8924	6669	|gi|1513069	|P-type adenosine triphosphatase |Listeria monocytogenes|	74	53	2256
55	11	11964	11401	|gnl|PID|e283110	|femD |Staphylococcus aureus|	74	64	564
61	2	1782	427	|gi|2293216	|(AF008220) putative UDP-N-acetylmuramate-alanine ligase |Bacillus subtilis|	74	55	1356
76	10	9414	8065	|gnl|PID|d101325	|YqiB |Bacillus subtilis|	74	54	1350
83	2	666	926	|pir|C33496|C334	|hisC homolog - Bacillus subtilis	74	55	261
86	9	8985	8080	|gi|683585	|prephenate dehydratase |Lactococcus lactis|	74	55	906

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
102	5	5005	5652	gi|143394	OMP-PRPP transferase (Bacillus subtilis)	74	57	648
103	5	4364	3267	gnl|PID|e323524	YloN protein (Bacillus subtilis)	74	62	1098
108	7	6864	7592	gnl|PID|e257631	methyltransferase (Lactococcus lactis)	74	56	729
131	2	478	146	gnl|PID|d101320	Yqg2 (Bacillus subtilis)	74	45	333
133	2	1380	919	gnl|PID|e313025	hypothetical protein (Bacillus subtilis)	74	60	462
137	9	6167	6787	gnl|PID|d100479	Na+ -ATPase subunit D (Enterococcus hirae)	74	53	621
149	4	3008	3883	gnl|PID|d100581	high level kasgamycin resistance (Bacillus subtilis)	74	55	876
157	2	243	824	gi|1573373	methylated-DNA--protein-cysteine methyltransferase (datI) (Haemophilus influenzae)	74	48	582
164	6	3515	4249	gi|410131	ORFX7 (Bacillus subtilis)	74	48	735
167	7	5446	5201	gi|413927	ipa-3r gene product (Bacillus subtilis)	74	55	246
171	1	1	1818	gnl|PID|d102251	beta-galactosidase (Bacillus circulans)	74	62	1818
172	4	1064	2392	gi|466474	cellobiose phosphotransferase enzyme II'' (Bacillus stearothermophilus)	74	50	1329
185	1	326	1	gi|1573646	Mg(2+) transport ATPase protein C (mgtC) (SP:P22037) (Haemophilus influenzae)	74	68	324
188	2	1089	2018	gi|1573008	ATP dependent translocator homolog (msbA) (Haemophilus influenzae)	74	44	930
189	11	6491	7174	gi|1661199	sakacin A production response regulator (Streptococcus mutans)	74	60	684
210	2	520	1287	gi|2293207	(AF008220) YtmQ (Bacillus subtilis)	74	60	768
261	1	836	192	gi|666983	putative ATP binding subunit (Bacillus subtilis)	74	55	645
263	3	1619	3655	gi|663232	Similarity with S. cerevisiae hypothetical 137.7 kD protein in subtelomeric Y' repeat region (Saccharomyces cerevisiae)	74	42	2037
265	2	844	1227	gi|49272	Asparaginase (Bacillus licheniformis)	74	64	384
368	1	1	942	gi|603998	unknown (Saccharomyces cerevisiae)	74	39	942
7	16	13357	11921	gnl|PID|d101324	YqhX (Bacillus subtilis)	73	57	1437
17	10	5706	5449	gnl|PID|e305362	unnamed protein product (Streptococcus thermophilus)	73	47	258
31	2	522	244	gnl|PID|d100576	single strand DNA binding protein (Bacillus subtilis)	73	55	279
32	6	5667	6194	gnl|PID|d101315	YqfG (Bacillus subtilis)	73	58	528
34	15	10281	9790	gnl|PID|d102151	(AB001684) ORF42c (Chlorella vulgaris)	73	46	492

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
40	12	9876	9226	gi|1173517	riboflavin synthase alpha subunit (Actinobacillus pleuropneumoniae)	73	55	651
55	2	3592	839	gnl|PID|d101887	cation-transporting ATPase PacL (Synechocystis sp.)	73	60	2754
55	18	17494	16586	gnl|PID|e265580	unknown (Mycobacterium tuberculosis)	73	52	909
65	16	7213	7767	gi|143419	ribosomal protein L6 (Bacillus stearothermophilus)	73	60	555
66	3	3300	3659	gnl|PID|e269883	LacF (Lactobacillus casei)	73	52	360
70	10	5557	5733	gi|857631	envelope protein (Human immunodeficiency virus type 1)	73	60	177
71	4	6133	8262	gnl|PID|e322063	ss-1,4-galactosyltransferase (Streptococcus pneumoniae)	73	45	2130
72	1	3	851	gi|2293177	(AF008220) transporter (Bacillus subtilis)	73	50	849
76	7	7019	6195	gnl|PID|d101325	YqiF (Bacillus subtilis)	73	66	825
76	12	10009	9533	gi|1573086	uridine kinase (uridine monophosphokinase) (udk) (Haemophilus influenzae)	73	54	477
80	7	8113	9372	gi|1377823	aminopeptidase (Bacillus subtilis)	73	60	1260
97	5	3389	1668	gnl|PID|d101954	dihydroxyacid dehydratase (Synechocystis sp.)	73	54	1722
98	9	6912	7619	gnl|PID|e314991	FtsE (Mycobacterium tuberculosis)	73	54	708
108	11	10928	10440	gi|388109	regulatory protein (Enterococcus faecalis)	73	54	489
128	6	3632	4222	gi|1685111	orf1091 (Streptococcus thermophilus)	73	63	591
138	2	1575	394	gi|147326	transport protein (Escherichia coli)	73	60	1182
140	13	12538	11903	pir|E53402|E534	serine O-acetyltransferase (EC 2.3.1.30) - Bacillus stearothermophilus	73	55	636
162	5	5701	4991	gnl|PID|e323511	putative YhaQ protein (Bacillus subtilis)	73	50	711
164	4	2323	2790	gi|1592076	hypothetical protein (SP:P25768) (Methanococcus jannaschii)	73	52	468
164	8	4815	5546	gi|410137	ORFX13 (Bacillus subtilis)	73	56	732
170	5	4394	5102	gnl|PID|d100959	homologue of unidentified protein of E. coli (Bacillus subtilis)	73	46	909
178	7	3893	4855	gi|46242	nodulation protein B, 5'end (Rhizobium loti)	73	56	963
204	6	5096	4278	gnl|PID|e214719	PlcR protein (Bacillus thuringiensis)	73	41	819
213	2	832	2037	gi|1565296	ribosomal protein S1 homolog; sequence specific DNA-binding protein (Leuconostoc lactis)	73	55	1206
231	2	84	287	gi|40173	homolog of E.coli ribosomal protein L21 (Bacillus subtilis)	73	61	204
237	1	2	505	gi|1773151	adenine phosphoribosyltransferase (Escherichia coli)	73	51	504

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
269	1	2	691	gnl|PID|d101328	YqiX (Bacillus subtilis)	73	36	690
289	2	1272	832	pir|A02771|R7MC	ribosomal protein L7/L12 - Micrococcus luteus	73	66	441
343	1	14	484	gi|1788125	(AE000276) hypothetical 30.4 kD protein in manZ-cspC intergenic region (Escherichia coli)	73	47	471
356	1	222	4	gi|2149905	D-glutamic acid adding enzyme (Enterococcus faecalis)	73	50	219
7	5	3165	4691	gnl|PID|d101833	amidase (Synechocystis sp.)	72	52	1527
7	9	7195	7647	gi|146976	nusB (Escherichia coli)	72	54	453
7	17	13743	13300	gnl|PID|e289141	similar to hydroxymyristoyl-(acyl carrier protein) dehydratase (Bacillus subtilis)	72	59	444
22	19	15637	16224	gnl|PID|d101929	ribosome releasing factor (Synechocystis sp.)	72	51	588
33	17	12111	11425	gnl|PID|d101190	ORF3 (Streptococcus mutans)	72	55	687
34	7	7147	5627	gi|396501	aspartyl-tRNA synthetase (Thermus thermophilus)	72	52	1521
38	23	15372	16085	pir|H64108|H641	L-ribulose-phosphate 4-epimerase (araD) homolog - Haemophilus influenzae (strain Rd KW20)	72	54	714
39	5	5094	6905	gnl|PID|e254877	unknown (Mycobacterium tuberculosis)	72	56	1812
40	6	4469	4636	gi|153672	lactose repressor (Streptococcus mutans)	72	58	168
48	2	1459	1253	gi|310380	inhibin beta-A-subunit (Ovis aries)	72	33	207
48	29	21729	22424	gi|2314329	(AE000623) glutamine ABC transporter, permease protein (glnP) (Helicobacter pylori)	72	49	696
50	5	4529	3288	gi|1750108	YnbA (Bacillus subtilis)	72	54	1242
51	3	1044	2282	gi|2293230	(AF008220) YtbJ (Bacillus subtilis)	72	54	1239
52	13	13681	13938	gi|142521	deoxyribodipyrimidine photolyase (Bacillus subtilis)	72	45	258
55	1	841	35	gi|882518	ORF_o304; GTG start (Escherichia coli)	72	59	807
75	5	2832	3191	gnl|PID|e209886	mercuric resistance operon regulatory protein (Bacillus subtilis)	72	44	360
76	6	6229	5771	gi|142450	ahrC protein (Bacillus subtilis)	72	53	459
79	5	5065	4592	gi|2293279	(AF008220) YtcG (Bacillus subtilis)	72	46	474
87	14	14726	12309	gnl|PID|e323502	putative PriA protein (Bacillus subtilis)	72	52	2418
91	1	444	662	gi|500691	MYO1 gene product (Saccharomyces cerevisiae)	72	50	219
91	7	4516	4764	gi|829615	skeletal muscle sodium channel alpha-subunit (Equus caballus)	72	38	249

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
95	2	2004	1717	gnl|PID|e323527	putative Asp23 protein (Bacillus subtilis)	72	40	288
109	1	1452	118	gi|143331	alkaline phosphatase regulatory protein (Bacillus subtilis)	72	52	1335
126	1	3	2192	gnl|PID|d101831	glutamine-binding periplasmic protein (Synechocystis sp.)	72	46	2190
130	3	1735	2478	gi|2415396	(AF015775) carboxypeptidase (Bacillus subtilis)	72	53	744
137	6	2585	2929	gi|472922	v-type Na-ATPase (Enterococcus hirae)	72	46	345
140	10	9601	9203	gi|49224	URF 4 (Synechococcus sp.)	72	48	399
146	5	1906	1247	gnl|PID|e324945	hypothetical protein (Bacillus subtilis)	72	45	660
147	2	2084	1083	gnl|PID|e325016	hypothetical protein (Bacillus subtilis)	72	56	1002
147	5	6156	5146	gi|472327	TPP-dependent acetoin dehydrogenase beta-subunit (Clostridium magnum)	72	56	1011
148	8	5381	6433	gi|974332	NAD(P)H-dependent dihydroxyacetone-phosphate reductase (Bacillus subtilis)	72	54	1053
148	14	10256	9675	gnl|PID|d101319	YqgN (Bacillus subtilis)	72	50	582
159	8	4005	4949	gi|1788770	(AE000330) o463; 24 pct identical (44 gaps) to 338 residues from penicillin-binding protein 4*, PBPE_BACSU SW: P32959 (451 aa) (Escherichia coli)	72	43	945
172	10	9907	10620	gi|763387	unknown (Saccharomyces cerevisiae)	72	55	714
220	3	2862	3602	gi|1574175	hypothetical (Haemophilus influenzae)	72	50	741
267	1	3	449	gi|290513	f470 (Escherichia coli)	72	48	447
281	2	899	540	gnl|PID|d100964	homologue of aspartokinase 2 alpha and beta subunits LysC of B. subtilis (Bacillus subtilis)	72	45	360
290	1	1018	14	gi|474195	This ORF is homologous to a 40.0 kd hypothetical protein in the htrB 3' region from E. coli, Accession Number X61000 (Mycoplasma-like organism)	72	54	1005
300	1	63	587	gi|746399	transcription elongation factor (Escherichia coli)	72	50	525
316	1	1326	4	gi|158127	protein kinase C (Drosophila melanogaster)	72	40	1323
342	1	227	3	gnl|PID|d101164	unknown (Bacillus subtilis)	72	54	225
354	1	1	1005	gnl|PID|d102048	C. thermocellum beta-glucosidase; P26208 (985) (Bacillus subtilis)	72	52	1005
6	10	8134	10467	gnl|PID|e264229	unknown (Mycobacterium tuberculosis)	71	57	2334
7	20	16231	15464	gi|18046	3-oxoacyl-[acyl-carrier protein] reductase (Cuphea lanceolata)	71	52	768
15	1	1297	2	gnl|PID|d100571	replicative DNA helicase (Bacillus subtilis)	71	51	1296
15	4	4435	3869	gi|499384	orf189 (Bacillus subtilis)	71	47	567

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
18	6	5120	4218	gnl|PID|d101318	YqgG (Bacillus subtilis)	71	51	903
29	1	1	540	gi|1773142	similar to the 20.2kd protein in TETB-EXOA region of B. subtilis (Escherichia coli)	71	56	540
38	20	13327	13830	gi|537036	ORF_o158 (Escherichia coli)	71	48	504
51	12	15015	12676	gi|149528	dipeptidyl peptidase IV (Lactococcus lactis)	71	55	2340
5	23	21040	20585	gi|2343285	(AF015453) surface located protein (Lactobacillus rhamnosus)	71	58	456
60	2	705	265	gnl|PID|d101320	YqgZ (Bacillus subtilis)	71	44	441
71	18	24679	26226	gi|580920	rodD (gtaA) polypeptide (AA 1-673) (Bacillus subtilis)	71	44	1548
71	25	30587	30360	gi|606028	ORF_o414; Geneplot suggests frameshift near start but none found (Escherichia coli)	71	50	228
72	6	5239	6729	gi|580835	lysine decarboxylase (Bacillus subtilis)	71	48	1491
72	14	11991	12878	gi|624085	similar to rat beta-alanine synthetase encoded by GenBank Accession Number S27881; contains ATP/GTP binding motif (Paramecium bursaria Chlorella virus 1)	71	54	888
73	11	7269	7033	gi|1906594	PNI (Rattus norvegicus)	71	42	237
74	6	10385	8517	gi|1573733	prolyl-tRNA synthetase (proS) (Haemophilus influenzae)	71	52	1869
81	9	5772	6578	gi|147404	mannose permease subunit II-M-Man (Escherichia coli)	71	45	807
86	5	4602	3604	gnl|PID|e322063	ss-1,4-galactosyltransferase (Streptococcus pneumoniae)	71	53	999
105	4	3619	4707	gi|2323341	(AF014460) PepQ (Streptococcus mutans)	71	58	1089
106	13	13557	12955	gi|1519287	LemA (Listeria monocytogenes)	71	48	603
114	2	1029	1979	gi|310303	mosA (Rhizobium meliloti)	71	55	951
122	2	564	1205	gi|1649037	glutamine transport ATP-binding protein GLNQ (Salmonella typhimurium)	71	50	642
132	5	9018	7063	gnl|PID|d102049	H. influenzae hypothetical ABC transporter; P44808 (974) (Bacillus subtilis)	71	51	1956
140	1	1141	227	gi|1673788	(AE000015) Mycoplasma pneumoniae, fructose-bisphosphate aldolase; similar to Swiss-Prot Accession Number P13243, from B subtilis (Mycoplasma pneumoniae)	71	49	915
140	5	5635	4973	gnl|PID|d100964	homologue of hypothetical protein in a rapamycin synthesis gene cluster of Streptomyces hygroscopicus (Bacillus subtilis)	71	48	663
141	7	7369	7845	gnl|PID|d102005	(AB001488) FUNCTION UNKNOWN, SIMILAR PRODUCT IN E. COLI AND MYCOPLASMA PNEUMONIAE. (Bacillus subtilis)	71	51	477

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
193	1	1	165	gi|46912	ribosomal protein L13 (Staphylococcus carnosus)	71	59	165
194	3	2205	1594	gi|535351	CodY (Bacillus subtilis)	71	52	612
199	3	1510	1319	gi|2182574	(AE000090) Y4pE (Rhizobium sp. NGR234)	71	45	192
208	2	2616	3752	gi|1787378	(AE000213) hypothetical protein in purB 5' region (Escherichia coli)	71	57	1137
209	2	2022	1141	gi|41432	fepC gene product (Escherichia coli)	71	46	882
210	5	1911	3071	gi|49316	ORF2 gene product (Bacillus subtilis)	71	45	1161
210	6	3069	3386	gi|580900	ORF3 gene product (Bacillus subtilis)	71	48	318
212	2	3561	1381	gi|557567	ribonucleotide reductase R1 subunit (Mycobacterium tuberculosis)	71	53	2181
233	3	2003	2920	gnl|PID|d101320	YqgR (Bacillus subtilis)	71	50	918
244	1	13	1053	gnl|PID|d100964	homologue of aspartokinase 2 alpha and beta subunits LysC of B. subtilis (Bacillus subtilis)	71	55	1041
251	2	1008	1874	gi|755601	unknown (Bacillus subtilis)	71	46	867
282	2	906	712	gi|1353874	unknown (Rhodobacter capsulatus)	71	46	195
312	4	2137	1565	gnl|PID|d102245	(AB005554) yxbF (Bacillus subtilis)	71	34	573
338	1	3	683	gi|1591045	hypothetical protein (SP:P31466) (Methanococcus jannaschii)	71	48	681
346	1	3	164	gi|1591234	hypothetical protein (SP:P42297) (Methanococcus jannaschii)	71	36	162
374	1	619	2	gi|397526	clumping factor (Staphylococcus aureus)	71	23	618
377	1	688	2	gi|397526	clumping factor (Staphylococcus aureus)	71	23	687
3	8	7419	6958	gnl|PID|e269486	Unknown (Bacillus subtilis)	70	42	462
3	10	8395	9075	gnl|PID|e255543	putative iron dependant repressor (Staphylococcus epidermidis)	70	46	681
7	14	11024	10254	gnl|PID|d100290	undefined open reading frame (Bacillus stearothermophilus)	70	55	771
7	18	14213	13719	gnl|PID|d101090	biotin carboxyl carrier protein of acetyl-CoA carboxylase (Synechocystis sp.)	70	56	495
9	2	1057	287	gnl|PID|d100581	unknown (Bacillus subtilis)	70	52	771
12	4	2610	1789	gnl|PID|d101195	yycJ (Bacillus subtilis)	70	52	822
21	2	2586	1846	gi|2293447	(AF008930) ATPase (Bacillus subtilis)	70	54	741
22	13	10955	11512	gi|1165295	Ydr540cp (Saccharomyces cerevisiae)	70	50	558
30	6	4315	3980	gi|39478	ATP binding protein of transport ATPases (Bacillus firmus)	70	51	336

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
31	1	370	113	gi|662792	single-stranded DNA binding protein (unidentified eubacterium)	70	36	258
33	15	10639	9521	gi|1161219	homolgous to D-amino acid dehydrogenase enzyme (Pseudomonas aeruginosa)	70	50	1119
38	6	3812	4312	gi|2058547	ComYD (Streptococcus gordonii)	70	48	501
38	25	17986	18477	gi|537033	ORF_f356 (Escherichia coli)	70	58	492
40	13	11054	9846	gi|1173516	riboflavin-specific deaminase (Actinobacillus pleuropneumoniae)	70	52	1209
42	2	722	1954	gi|1146183	putative (Bacillus subtilis)	70	51	1233
43	3	2373	1612	gi|1591493	glutamine transport ATP-binding protein Q (Methanococcus jannaschii)	70	48	762
45	8	9197	8049	gnl|PID|d102036	subunit of ADP-glucose pyrophosphorylase (Bacillus stearothermophilus)	70	54	1149
59	2	567	956	gnl|PID|d100302	neopullulanase (Bacillus sp.)	70	42	390
60	3	1874	795	gnl|PID|e276466	aminopeptidase P (Lactococcus lactis)	70	48	1080
61	4	5553	2437	gnl|PID|e275074	SNF (Bacillus cereus)	70	51	3117
61	7	7914	6802	gi|1573037	cystathionine gamma-synthase (metB) (Haemophilus influenzae)	70	52	1113
63	7	5372	7222	gnl|PID|d100974	unknown (Bacillus subtilis)	70	54	1851
68	7	7126	6962	gi|1263014	emm18.1 gene product (Streptococcus pyogenes)	70	37	165
72	12	10081	10911	gi|2313093	(AE000524) carboxynorspermidine decarboxylase (nspC) (Helicobacter pylori)	70	56	831
75	10	7888	8124	gi|1877423	galactose-1-P-uridyl transferase (Streptococcus mutans)	70	59	237
79	3	3424	2525	gi|39881	ORF 311 (AA 1-311) (Bacillus subtilis)	70	47	900
87	10	9369	7324	gnl|PID|e323506	putative Pkn2 protein (Bacillus subtilis)	70	52	2046
96	14	10640	11788	gi|1573209	tRNA-guanine transglycosylase (tgt) (Haemophilus influenzae)	70	52	1149
113	2	574	1086	gi|433630	A180 (Saccharomyces cerevisiae)	70	59	513
123	5	2901	3461	gnl|PID|d100585	unknown (Bacillus subtilis)	70	45	561
125	5	4593	4282	gnl|PID|e276474	capacitative calcium entry channel 1 (Bos taurus)	70	35	312
129	5	4500	3454	gnl|PID|d101314	YqeT (Bacillus subtilis)	70	47	1047
133	1	2608	1394	gi|2293312	(AF008220) YtfP (Bacillus subtilis)	70	50	1215
135	1	420	662	gnl|PID|e265530	yorfE (Streptococcus pneumoniae)	70	47	243
137	3	438	932	gi|472919	v-type Na-ATPase (Enterococcus hirae)	70	57	495
138	1	440	3	gi|147336	transmembrane protein (Escherichia coli)	70	42	438

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
140	16	18796	16364	gi|976441	N5-methyltetrahydrofolate homocysteine methyltransferase (Saccharomyces cerevisiae)	70	53	2433
167	10	8263	6695	gi|149535	D-alanine activating enzyme (Lactobacillus casei)	70	52	1569
204	4	3226	2747	gnl|PID|d102049	E. coli hypothetical protein; P31805 (267) (Bacillus subtilis)	70	51	480
207	3	2627	2869	gnl|PID|e309213	racGAP (Dictyostelium discoideum)	70	45	243
282	3	1136	882	gi|1353874	unknown (Rhodobacter capsulatus)	70	50	255
6	21	17554	18453	gnl|PID|e233879	hypothetical protein (Bacillus subtilis)	69	44	900
6	22	18482	19471	gi|580883	ipa-88d gene product (Bacillus subtilis)	69	53	990
22	6	4682	5824	gi|2209379	(AF006720) ProJ (Bacillus subtilis)	69	48	1143
22	9	7992	8651	gnl|PID|d100580	unknown (Bacillus subtilis)	69	51	660
22	12	9871	10767	gnl|PID|d100581	unknown (Bacillus subtilis)	69	51	897
27	7	5857	5348	gnl|PID|d102012	(AB001488) FUNCTION UNKNOWN. (Bacillus subtilis)	69	28	510
36	10	7294	10116	gi|437916	isoleucyl-tRNA synthetase (Staphylococcus aureus)	69	53	2823
38	1	2	1090	gi|141900	alcohol dehydrogenase (EC 1.1.1.1) (Alcaligenes eutrophus)	69	48	1089
40	14	11333	11944	gi|1573280	Holliday junction DNA helicase (ruvA) (Haemophilus influenzae)	69	44	612
40	15	11942	12517	gi|1573653	DNA-3-methyladenine glycosidase I (tagI) (Haemophilus influenzae)	69	50	576
45	6	6947	5490	gi|580887	starch (bacterial glycogen) synthase (Bacillus subtilis)	69	47	1458
48	14	24932	24153	gnl|PID|e233870	hypothetical protein (Bacillus subtilis)	69	36	780
49	6	6183	6521	gi|396297	similar to phosphotransferase system enzyme II (Escherichia coli)	69	50	339
49	8	7586	8338	gi|396420	similar to Alcaligenes eutrophus pHG1 D-ribulose-5-phosphate 3 epimerase (Escherichia coli)	69	49	753
55	6	8262	7033	gi|1146238	poly(A) polymerase (Bacillus subtilis)	69	50	1230
59	3	954	2333	gnl|PID|e313038	hypothetical protein (Bacillus subtilis)	69	54	1380
62	3	1170	1418	gnl|PID|d101915	hypothetical protein (Synechocystis sp.)	69	49	249
63	8	7298	7762	gi|293017	ORF3 (put.); putative (Lactococcus lactis)	69	42	465
66	4	3657	5081	gi|153755	phospho-beta-D-galactosidase (EC 3.2.1.85) (Lactococcus lactis cremoris)	69	49	1425
66	5	5126	6829	gi|433809	enzyme II (Streptococcus mutans)	69	46	1704
71	6	10017	10664	gnl|PID|e322063	ss-1,4-galactosyltransferase (Streptococcus pneumoniae)	69	39	648

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
71	21	27730	27966	gnl|PID|d100649	DE-cadherin (Drosophila melanogaster)	69	30	237
77	1	1	237	gi|287870	groES gene product (Lactococcus lactis)	69	44	237
81	5	3622	4101	gi|1573605	fucose operon protein (fucU) (Haemophilus influenzae)	69	52	480
83	1	40	714	pir|C33496|C334	hisC homolog - Bacillus subtilis	69	46	675
83	16	15742	16335	gi|143372	phosphoribosyl glycinamide formyltransferase (PUR-N) (Bacillus subtilis)	69	46	594
85	2	1212	916	gi|194097	IFN-response element binding factor 1 (Mus musculus)	69	48	297
91	5	3678	4274	gi|1574712	anaerobic ribonuleoside-triphosphate reductase activating protein (nrdG) (Haemophilus influenzae)	69	44	597
98	5	3247	4032	gnl|PID|d100262	LivF protein (Salmonella typhimurium)	69	51	786
108	5	4085	5056	gnl|PID|e257629	transcription factor (Lactococcus lactis)	69	49	972
126	3	3078	4568	gnl|PID|d101329	YqjJ (Bacillus subtilis)	69	49	1491
131	6	4121	2889	gnl|PID|d101314	YqeR (Bacillus subtilis)	69	47	1233
136	2	1505	2299	gnl|PID|d100581	unknown (Bacillus subtilis)	69	47	795
149	5	3852	4763	gnl|PID|e323525	YloQ protein (Bacillus subtilis)	69	50	912
149	12	9336	10655	gi|151571	Homology with E.coli and P.aeruginosa lysA gene; product of unknown function; putative (Pseudomonas syringae)	69	52	1320
153	4	3191	3829	gi|1710373	BrnQ (Bacillus subtilis)	69	44	639
169	3	849	2124	gnl|PID|d100582	temperature sensitive cell division (Bacillus subtilis)	69	49	1476
180	1	566	3	gi|488339	alpha-amylase (unidentified cloning vector)	69	50	564
212	1	1196	231	gi|1395209	ribonucleotide reductase R2-2 small subunit (Mycobacterium tuberculosis)	69	53	966
226	1	2	661	pir|JQ2285|JQ22	nodulin-26 - soybean	69	41	660
233	5	3249	4766	gi|472918	v-type Na-ATPase (Enterococcus hirae)	69	56	1518
235	3	660	1766	gi|148945	methylase (Haemophilus influenzae)	69	43	1107
243	2	865	2361	gnl|PID|d100225	ORF5 (Barley yellow dwarf virus)	69	69	1497
251	3	2899	1967	gi|2289231	macrolide-efflux protein (Streptococcus agalactiae)	69	51	933
310	1	1	282	gnl|PID|e322442	peptide deformylase (Clostridium beijerinckii)	69	55	282
369	1	868	2	gi|397526	clumping factor (Staphylococcus aureus)	69	22	867
370	1	749	3	gi|397526	clumping factor (Staphylococcus aureus)	69	21	747

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
379	1	44	280	|gnl|PID|d100649	DE-cadherin (Drosophila melanogaster)	69	30	237
388	1	260	72	|gi|1787524	(AE000225) hypothetical 32.7 kD protein in trpL-btuR intergenic region (Escherichia coli)	69	44	189
1	2	2006	3040	|gnl|PID|d101809	ABC transporter (Synechocystis sp.)	68	43	1035
12	5	3958	2600	|gi|2182992	histidine kinase (Lactococcus lactis cremoris)	68	45	1359
15	2	1790	1311	|pir|S16974|R5BS	ribosomal protein L9 - Bacillus stearothermophilus	68	56	480
16	6	7353	5701	|gi|1787041	(AE000184) o530; This 530 aa ort is 33 pct identical (14 gaps) to 525 residues of an approx. 640 aa protein YHES_HAEIN SW: P44808 (Escherichia coli)	68	45	1653
17	12	6479	6805	|gi|553165	acetylcholinesterase (Homo sapiens)	68	68	327
20	13	14128	14505	|gi|142700	P competence protein (ttg start codon) (put.); putative (Bacillus subtilis)	68	40	378
22	32	24612	25397	|gi|289262	comE ORF3 (Bacillus subtilis)	68	36	786
30	7	4548	4288	|gi|311388	ORF1 (Azorhizobium caulinodans)	68	46	261
36	5	3911	4585	|gi|1573041	hypothetical (Haemophilus influenzae)	68	54	675
46	6	5219	6040	|gi|1790131	(AE000446) hypothetical 29.7 kD protein in ibpA-gyrB intergenic region (Escherichia coli)	68	47	822
54	10	6235	7086	|gi|882579	CG Site No. 29739 (Escherichia coli)	68	55	852
55	5	7069	5165	|gnl|PID|d101914	ABC transporter (Synechocystis sp.)	68	45	1905
71	3	6134	5613	|gi|1573353	outer membrane integrity protein (tolA) (Haemophilus influenzae)	68	50	522
71	10	15342	16613	|gi|580866	ipa-12d gene product (Bacillus subtilis)	68	31	1272
71	12	17560	18792	|gi|44073	SecY protein (Lactococcus lactis)	68	35	1233
71	17	22295	24703	|gi|1762349	involved in protein export (Bacillus subtilis)	68	50	2409
73	16	10208	9729	|gi|1353537	dUTPase (Bacteriophage r1t)	68	51	480
86	18	17198	16011	|gi|413943	ipa-19d gene product (Bacillus subtilis)	68	53	1188
87	17	17491	15866	|gi|150209	ORF 1 (Mycoplasma mycoides)	68	43	1626
89	6	5139	4354	|gi|1498824	M. jannaschii predicted coding region MJ0062 (Methanococcus jannaschii)	68	40	786
89	11	8021	8242	|gi|150974	4-oxalocrotonate tautomerase (Pseudomonas putida)	68	43	222
97	8	6755	5394	|gi|2367358	(AE000491) hypothetical 52.9 kD protein in aidB-rpsF intergenic region (Escherichia coli)	68	41	1362

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
98	3	1418	2308	gnl|PID|d100261	LivA protein (Salmonella typhimurium)	68	40	891
99	13	16414	17280	gi|455363	regulatory protein (Streptococcus mutans)	68	50	867
115	3	5054	3693	gi|466474	cellobiose phosphotransferase enzyme II'' (Bacillus stearothermophilus)	68	44	1362
124	7	3394	3221	gnl|PID|d100702	cut14 protein (Schizosaccharomyces pombe)	68	56	174
125	2	2923	1922	gi|450566	transmembrane protein (Bacillus subtilis)	68	50	1002
132	2	4858	2888	gnl|PID|d101732	DNA ligase (Synechocystis sp.)	68	52	1971
140	7	7765	7580	gi|1209711	unknown (Saccharomyces cerevisiae)	68	47	186
150	1	539	3	gi|402490	ADP-ribosylarginine hydrolase (Mus musculus)	68	59	537
164	1	58	867	gnl|PID|e255114	glutamate racemase (Bacillus subtilis)	68	49	810
	2	819	1835	gnl|PID|e255117	hypothetical protein (Bacillus subtilis)	68	50	1017
169	7	3946	4104	pir|BS4545|BS45	hypothetical protein - Lactococcus lactis subsp. lactis plasmid pSL2	68	40	159
170	4	4247	4396	gi|304146	spore coat protein (Bacillus subtilis)	68	52	150
171	8	6002	7054	gi|38722	precursor (aa -20 to 381) (Acinetobacter calcoaceticus)	68	54	1053
198	3	2473	1871	gnl|PID|e313075	hypothetical protein (Bacillus subtilis)	68	46	603
211	2	969	1802	gi|1439528	EIIC-man (Lactobacillus curvatus)	68	45	834
214	8	4926	4231	gnl|PID|d102049	H. influenzae hypothetical protein; P43990 (182) (Bacillus subtilis)	68	50	696
217	6	4955	5170	gnl|PID|e326966	similar to B.vulgaris CMS-associated mitochondrial ... (reverse transcriptase) (Arabidopsis thaliana)	68	36	216
218	7	3930	4745	gi|2293198	(AF008220) YtgP (Bacillus subtilis)	68	38	816
220	6	4628	4338	gnl|PID|e325791	(AJ000005) orf1 (Bacillus megaterium)	68	51	291
236	1	746	108	gi|410137	ORFX13 (Bacillus subtilis)	68	46	639
237	2	675	1451	gi|396348	homoserine transsuccinylase (Escherichia coli)	68	49	777
250	4	771	1229	gi|310859	ORF2 (Synechococcus sp.)	68	50	459
254	1	517	155	gi|1787105	(AE000189) o648 was o669; This 669 aa orf is 40 pct identical (1 gaps) to 217 residues of an approx. 232 aa protein YBBA_HAEIN SW: P45247 (Escherichia coli)	68	44	363
337	1	1	774	gnl|PID|e261990	putative orf (Bacillus subtilis)	68	47	774
345	1	3	653	gi|149513	thymidylate synthase (EC 2.1.1.45) (Lactococcus lactis)	68	61	651

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
386	2	417	4	|gi|1573353	|outer membrane integrity protein (tolA) (Haemophilus influenzae)	68	51	414
2	4	5722	4697	|gi|1592141	|M. jannaschii predicted coding region MJ1507 (Methanococcus jannaschii)	67	26	1026
3	6	5397	4591	|gi|2293175	|(AF008220) signal transduction regulator (Bacillus subtilis)	67	44	807
5	2	2301	574	|gi|2313385	|(AE000547) para-aminobenzoate synthetase (pabB) (Helicobacter pylori)	67	48	1728
6	19	16063	16758	|gi|413931	|ipa-7d gene product (Bacillus subtilis)	67	41	696
22	8	7094	7897	|gi|1928962	|pyrroline-5-carboxylate reductase (Actinidia deliciosa)	67	51	804
29	10	8335	9072	|gi|468745	|gtcR gene product (Bacillus brevis)	67	41	738
31	3	1379	585	|gi|2425123	|(AF019986) PksB (Dictyostelium discoideum)	67	49	795
32	11	8849	10150	|gi|42029	|ORF1 gene product (Escherichia coli)	67	47	1302
36	16	14830	15546	|gi|1592142	|ABC transporter, probable ATP-binding subunit (Methanococcus jannaschii)	67	43	717
38	9	4958	5392	|gnl|PID|e214803	|T22B3.3 (Caenorhabditis elegans)	67	47	435
38	21	13775	14512	|gi|537037	|ORF_o216 (Escherichia coli)	67	52	738
45	9	10428	9181	|gi|551710	|branching enzyme (glgB) (EC 2.4.1.18) (Bacillus stearothermophilus)	67	51	1248
48	23	18344	17514	|gi|413949	|ipa-25d gene product (Bacillus subtilis)	67	50	831
50	2	1773	952	|gnl|PID|d101330	|YqjQ (Bacillus subtilis)	67	55	822
53	1	431	3	|gi|1574291	|fimbrial transcription regulation repressor (pilB) (Haemophilus influenzae)	67	40	429
55	13	12740	11946	|gnl|PID|e252990	|ORF YDL037c (Saccharomyces cerevisiae)	67	51	795
61	9	9210	8329	|gnl|PID|e264711	|ATP-binding cassette transporter A (Staphylococcus aureus)	67	50	882
71	2	5614	6117	|gi|1197667	|vitellogenin (Anolis pulchellus)	67	36	504
81	7	4489	4983	|gi|1142714	|phosphoenolpyruvate:mannose phosphotransferase element IIB (Lactobacillus curvatus)	67	42	495
83	7	2957	3214	|gi|1276746	|Acyl carrier protein (Porphyra purpurea)	67	37	258
86	8	8140	6809	|gi|1147744	|PSR (Enterococcus hirae)	67	45	1332
97	3	986	1366	|gnl|PID|d102235	|(AB000631) unnamed protein product (Streptococcus mutans)	67	43	381
102	1	601	1413	|gi|682765	|mccB gene product (Escherichia coli)	67	36	813
106	3	1109	1987	|gi|148921	|LicD protein (Haemophilus influenzae)	67	43	879
115	4	5982	5656	|gi|895750	|putative cellobiose phosphotransferase enzyme III (Bacillus subtilis)	67	44	327

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
115	7	8421	8077	|gi|466473	|cellobiose phosphotransferase enzyme II' (Bacillus stearothermophilus)	67	51	345
127	13	8127	7021	|gi|147326	|transport protein (Escherichia coli)	67	45	1107
136	3	2215	2859	|gnl|PID|d100581	|unknown (Bacillus subtilis)	67	49	645
140	21	23317	20906	|gnl|PID|d101912	|phenylalanyl-tRNA synthetase (Synechocystis sp.)	67	43	2412
146	6	2894	1893	|gi|2182994	|histidine kinase (Lactococcus lactis cremoris)	67	44	1002
151	8	11476	11117	|gnl|PID|d100085	|ORF129 (Bacillus cereus)	67	48	360
160	10	7453	8646	|gi|2281317	|OrfB; similar to a Streptococcus pneumoniae putative membrane protein encoded by GenBank Accession Number X99400; inactivation of the OrfB gene leads to UV-sensitivity and to decrease of homologous recombination (plasmidic test) (Lactococcus l	67	46	1194
163	3	3099	4505	|gnl|PID|d101317	|YqfR (Bacillus subtilis)	67	47	1407
167	8	6704	5454	|gi|1161933	|DltB (Lactobacillus casei)	67	45	1251
169	4	2322	2879	|gnl|PID|d101331	|YqkG (Bacillus subtilis)	67	41	558
171	11	7656	8384	|gi|153841	|pneumococcal surface protein A (Streptococcus pneumoniae)	67	50	729
188	3	1930	3723	|gi|1542975	|AbcB (Thermoanaerobacterium thermosulfurigenes)	67	46	1794
189	6	3599	3141	|gnl|PID|e325178	|Hypothetical protein (Bacillus subtilis)	67	52	459
205	3	1663	2211	|gi|606073	|ORF_o169 (Escherichia coli)	67	47	549
207	4	2896	3456	|gi|2276374	|DtxR/iron regulated lipoprotein precursor (Corynebacterium diphtheriae)	67	49	561
217	3	4086	3703	|gi|895750	|putative cellobiose phosphotransferase enzyme III (Bacillus subtilis)	67	42	384
246	2	291	662	|gi|1842438	|unknown (Bacillus subtilis)	67	43	372
252	1	2	745	|gi|2351768	|PspA (Streptococcus pneumoniae)	67	41	744
265	3	1134	1811	|gi|2313847	|(AE000585) L-asparaginase II (ansB) (Helicobacter pylori)	67	42	678
295	1	1	375	|gi|2276374	|DtxR/iron regulated lipoprotein precursor (Corynebacterium diphtheriae)	67	43	375
1	7	4898	5146	|gnl|PID|e255179	|unknown (Mycobacterium tuberculosis)	66	56	249
3	1	389	3	|gnl|PID|e269548	|Unknown (Bacillus subtilis)	66	48	387
3	20	19267	20805	|gi|39956	|IIGlc (Bacillus subtilis)	66	50	1539
4	3	2545	2718	|gi|1787564	|(AE000228) phage shock protein C (Escherichia coli)	66	36	174
5	9	13197	12592	|gi|1574291	|fimbrial transcription regulation repressor (pilB) (Haemophilus influenzae)	66	46	606

TABLE 2 S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
9	4	2872	1451	gnl|PID|e266928	unknown (Mycobacterium tuberculosis)	66	43	1422
12	2	1469	1200	gi|520407	orf2; GTG start codon (Bacillus thuringiensis)	66	42	270
15	12	10979	9897	gi|2314738	(AE000653) translation elongation factor EF-Ts (tsf) (Helicobacter pylori)	66	49	1083
16	2	1312	734	gnl|PID|d102245	(AB005554) yxbF (Bacillus subtilis)	66	35	579
22	3	1372	1851	gi|1480916	signal peptidase type II (Lactococcus lactis)	66	38	480
22	7	5828	7096	gnl|PID|e206261	gamma-glutamyl phosphate reductase (Streptococcus thermophilus)	66	51	1269
22	20	16194	17138	gnl|PID|e281914	YitL (Bacillus subtilis)	66	50	945
30	2	530	976	gi|2314379	(AE000627) ABC transporter, ATP-binding protein (yhcG) (Helicobacter pylori)	66	40	447
32	1	199	984	gi|312444	ORF2 (Bacillus caldolyticus)	66	49	786
33	13	8352	7234	gi|1387979	44% identity over 302 residues with hypothetical protein from Synechocystis sp. accession D64006_CD; expression induced by environmental stress; some similarity to glycosyl transferases; two potential membrane-spanning helices (Bacillus subtil	66	44	1119
34	6	5658	4708	gnl|PID|e250724	orf2 (Lactobacillus sake)	66	39	951
34	14	9792	9574	gi|1590997	M. jannaschii predicted coding region MJ0272 (Methanococcus jannaschii)	66	48	219
35	16	15163	14501	gi|1771352	Cap5M (Staphylococcus aureus)	66	46	663
36	9	6173	6976	gi|1518680	minicell-associated protein DivIVA (Bacillus subtilis)	66	35	804
36	11	10396	10824	bbs|155344	insulin activator factor, INSAF (human, Pancreatic insulinoma, Peptide Partial, 744 aa) (Homo sapiens)	66	43	429
48	1	28	1419	gnl|PID|e325204	hypothetical protein (Bacillus subtilis)	66	50	1392
48	7	3810	4112	gi|2182574	(AE000090) Y4pE (Rhizobium sp. NGR234)	66	40	303
52	4	3595	2789	gi|388565	major cell-binding factor (Campylobacter jejuni)	66	52	807
54	3	2662	1076	gnl|PID|d101831	glutamine-binding periplasmic protein (Synechocystis sp.)	66	43	1587
61	10	9740	9183	gnl|PID|e154144	mdr gene product (Staphylococcus aureus)	66	44	558
72	13	10893	11993	gi|2313129	(AE000526) H. pylori predicted coding region HP0049 (Helicobacter pylori)	66	44	1101
74	9	13267	12476	gi|1573941	hypothetical (Haemophilus influenzae)	66	43	792
75	1	2	868	gi|1574631	nicotinamide mononucleotide transporter (pnuC) (Haemophilus influenzae)	66	48	867
75	7	5303	4275	gi|41312	put. EBG repressor protein (Escherichia coli)	66	40	1029

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
82	7	6813	8123	gnl|PID|e255128	trigger factor (Bacillus subtilis)	66	53	1311
83	3	905	1219	pir|C33496|C334	hisC homolog - Bacillus subtilis	66	44	315
86	10	9407	8925	gi|683584	shikimate kinase (Lactococcus lactis)	66	41	483
88	10	7001	6060	gi|2098719	putative fimbrial-associated protein (Actinomyces naeslundii)	66	52	942
89	1	951	4	gi|410118	ORFX19 (Bacillus subtilis)	66	41	948
93	7	3661	2711	gi|1787936	(AE000260) f298; This 298 aa orf is 51 pct identical (5 gaps) to 297 residues of an approx. 304 aa protein YCSH_BACSU SW: Q42972 (Escherichia coli)	66	49	951
104	3	1805	3049	gi|1469784	putative cell division protein ftsW (Enterococcus hirae)	66	48	1245
106	14	13576	14253	gi|40027	homologous to E.coli gidB (Bacillus subtilis)	66	52	678
107	3	965	1864	gi|144858	ORF A (Clostridium perfringens)	66	49	900
112	7	5718	6593	gi|609332	DprA (Haemophilus influenzae)	66	43	876
115	1	3	302	gi|727367	Hyr1p (Saccharomyces cerevisiae)	66	56	300
122	1	3	566	gnl|PID|d101328	YqiY (Bacillus subtilis)	66	36	564
126	8	11759	11046	gnl|PID|d101163	ORF3 (Bacillus subtilis)	66	48	714
128	11	8201	8431	gi|726288	growth associated protein GAP-43 (Xenopus laevis)	66	41	231
131	8	4894	4508	gi|486661	THm1 related protein (Saccharomyces cerevisiae)	66	39	387
140	3	3236	2574	gi|40056	phoP gene product (Bacillus subtilis)	66	36	663
140	15	16318	15434	gi|1658189	5,10-methylenetetrahydrofolate reductase (Erwinia carotovora)	66	48	885
146	12	7926	7636	gnl|PID|d101140	transposase (Synechocystis sp.)	66	42	291
147	6	7137	6154	gi|472326	TPP-dependent acetoin dehydrogenase alpha-subunit (Clostridium magnum)	66	48	984
149	6	4435	5430	gnl|PID|d101887	pentose-5-phosphate-3-epimerase (Synechocystis sp.)	66	46	996
149	11	10754	11575	gi|42371	pyruvate formate-lyase activating enzyme (AA 1-246) (Escherichia coli)	66	42	822
186	4	2578	2270	gnl|PID|d101199	ORF11 (Enterococcus faecalis)	66	41	309
207	2	2340	2597	gnl|PID|e)21893	envelope glycoprotein gp160 (Human immunodeficiency virus type 1)	66	46	258
210	7	3358	3678	gi|49318	ORF4 gene product (Bacillus subtilis)	66	46	321
217	8	5143	5355	gi|49538	thrombin receptor (Cricetulus longicaudatus)	66	38	213
220	4	3875	3642	gi|466648	alternate name ORFD of L23635 (Escherichia coli)	66	33	234

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
221	1	1070	138	|gnl|PID|e247187	|zinc finger protein (Bacteriophage phigle)	66	45	933
224	2	1864	2640	|gi|1176199	|putative ABC transporter subunit (Staphylococcus epidermidis)	66	41	777
243	1	3	872	|dbj||AB000617_2	|(AB000617) YcdII (Bacillus subtilis)	66	45	870
268	2	891	568	|gi|517210	|putative transposase (Streptococcus pyogenes)	66	60	324
322	1	2	643	|gi|1499836	|Zn protease (Methanococcus jannaschii)	66	40	642
5	10	13909	13178	|gi|1574292	|hypothetical (Haemophilus influenzae)	65	34	732
6	11	10465	11190	|gi|142854	|homologous to E. coli radC gene product and to unidentified protein from Staphylococcus aureus (Bacillus subtilis)	65	48	726
7	2	647	405	|pir|C64146|C641	|hypothetical protein HI0259 - Haemophilus influenzae (strain Rd KW20)	65	42	243
7	7	6246	6821	|gnl|PID|d101323	|YqhU (Bacillus subtilis)	65	50	576
10	2	1873	1397	|gi|1163111	|ORF-1 (Streptococcus pneumoniae)	65	54	477
16	3	1428	2222	|gnl|PID|e325010	|hypothetical protein (Bacillus subtilis)	65	45	795
21	4	3815	3357	|gnl|PID|e314910	|hypothetical protein (Staphylococcus sciuri)	65	40	459
22	34	25776	26384	|gi|1123030	|CpxA (Actinobacillus pleuropneumoniae)	65	42	609
43	2	1648	290	|gi|1044826	|F14E5.1 (Caenorhabditis elegans)	65	38	1359
48	13	10062	10856	|gi|1573390	|hypothetical (Haemophilus influenzae)	65	45	795
48	22	17521	16883	|gi|1573391	|hypothetical (Haemophilus influenzae)	65	37	639
48	25	19027	18533	|gnl|PID|e264484	|YCR020c, len:215 (Saccharomyces cerevisiae)	65	38	495
49	3	3856	5334	|gi|1480429	|putative transcriptional regulator (Bacillus stearothermophilus)	65	32	1479
50	6	5337	4519	|gi|171963	|tRNA isopentenyl transferase (Saccharomyces cerevisiae)	65	42	819
52	15	14728	15588	|gi|1499745	|M. jannaschii predicted coding region MJ0912 (Methanococcus jannaschii)	65	46	861
59	7	3963	4745	|gi|496514	|orf zeta (Streptococcus pyogenes)	65	42	783
68	3	2500	3483	|gi|887824	|ORF_o310 (Escherichia coli)	65	46	984
69	3	2171	1077	|gnl|PID|e311453	|unknown (Bacillus subtilis)	65	42	1095
69	7	6029	5325	|gi|809660	|deoxyribose-phosphate aldolase (Bacillus subtilis)	65	55	705
71	5	8536	9783	|gi|1573224	|glycosyl transferase lgtC (GP:U14554_4) (Haemophilus influenzae)	65	42	1248
72	8	7664	8527	|gnl|PID|e267589	|Unknown, highly similar to several spermidine synthases (Bacillus subtilis)	65	39	864

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
76	5	5773	4097	gnl|PID|d101723	DNA REPAIR PROTEIN RECN (RECOMBINATION PROTEIN N). [Escherichia coli]	65	44	1677
76	9	8099	7875	gi|1574276	exodeoxyribonuclease, small subunit (xseB) [Haemophilus influenzae]	65	38	225
84	2	2870	2352	gi|2313188	(AE000532) conserved hypothetical protein [Helicobacter pylori]	65	41	519
86	15	14495	13407	gnl|PID|d101880	3-dehydroquinate synthase [Synechocystis sp.]	65	44	1089
87	3	3706	2423	gi|151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii]	65	51	1284
88	3	2425	2736	gi|1098510	unknown [Lactococcus lactis]	65	30	312
89	2	1627	1007	gnl|PID|d102008	(AB001488) SIMILAR TO ORF14 OF ENTEROCOCCUS FAECALIS TRANSPOSON TN916. [Bacillus subtilis]	65	41	621
111	6	6635	6186	gnl|PID|e246063	NM23/nucleoside diphosphate kinase [Xenopus laevis]	65	50	450
116	1	3	1016	gnl|PID|d101125	queuosine biosynthesis protein QueA [Synechocystis sp.]	65	44	1014
123	1	69	389	gi|498839	ORF2 [Clostridium perfringens]	65	36	321
123	7	6522	7190	gi|1575577	DNA-binding response regulator [Thermotoga maritima]	65	39	669
125	3	3821	2859	gnl|PID|e257609	sugar-binding transport protein [Anaerocellum thermophilum]	65	47	963
137	12	8015	7818	gi|2182574	(AE000090) Y4pE [Rhizobium sp. NGR234]	65	41	198
147	4	5021	3885	gi|472329	dihydrolipoamide acetyltransferase [Clostridium magnum]	65	47	1137
148	2	1053	1931	gnl|PID|d101319	YqgH [Bacillus subtilis]	65	42	879
151	2	3212	4687	gi|304897	EcoE type I restriction modification enzyme M subunit [Escherichia coli]	65	50	1476
156	2	730	437	gi|310893	membrane protein [Theileria parva]	65	47	294
164	7	4256	4817	gi|410132	ORFX8 [Bacillus subtilis]	65	48	582
169	6	3192	3914	gi|1552737	similar to purine nucleoside phosphorylase (deoD) [Escherichia coli]	65	41	723
176	4	2951	2220	gnl|PID|e339500	oligopeptide binding lipoprotein [Streptococcus pneumoniae]	65	43	732
195	4	4556	3900	gi|1592142	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	65	40	657
196	1	160	1572	gnl|PID|d102004	(AB001488) PROBABLE UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMYL-2, 6-DIAMINOLIGASE (EC 6.3.2.15). [Bacillus subtilis]	65	51	1413
204	2	2246	1215	gi|143156	membrane bound protein [Bacillus subtilis]	65	37	1032
210	4	1544	1891	gi|49315	ORF1 gene product [Bacillus subtilis]	65	48	348
242	2	1625	723	gi|1787540	(AE000226) f249; This 249 aa orf is 32 pct identical (8 gaps) to 244 residues of an approx. 272 aa protein AGAR_ECOLI SW: P42902 [Escherichia coli]	65	42	903

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
284	1	1	900	gi|559861	clyM (Plasmid pAD1)	65	36	900
304	1	2	574	gnl|PID|e290934	unknown (Mycobacterium tuberculosis)	65	52	573
315	1	2	1483	gi|790694	mannuronan C-5-epimerase (Azotobacter vinelandii)	65	57	1482
320	1	3	569	gnl|PID|d102048	K. aerogenes, histidine utilization repressor; P12380 (199) DNA binding (Bacillus subtilis)	65	46	567
358	1	1	309	gnl|PID|e323508	YloS protein (Bacillus subtilis)	65	55	309
2	7	7571	6696	gi|1498753	nicotinate-nucleotide pyrophosphorylase (Rhodospirillum rubrum)	64	47	876
6	6	5924	6802	gnl|PID|d101111	methionine aminopeptidase (Synechocystis sp.)	64	52	879
8	4	3417	3686	gi|1045935	DNA helicase II (Mycoplasma genitalium)	64	58	270
11	4	3249	2689	gnl|PID|e265529	OrfB (Streptococcus pneumoniae)	64	46	561
15	7	6504	7145	gi|1762328	Ycr59c/Yig2 homolog (Bacillus subtilis)	64	45	642
22	11	9548	9895	gnl|PID|d100581	unknown (Bacillus subtilis)	64	38	348
22	30	22503	23174	gi|289260	comE ORF1 (Bacillus subtilis)	64	44	672
26	7	14375	14199	gi|409286	bmrU (Bacillus subtilis)	64	30	177
27	2	1510	1334	gi|40795	DdeI methylase (Desulfovibrio vulgaris)	64	51	177
29	2	614	297	gi|2326168	type VII collagen (Mus musculus)	64	50	318
35	2	368	721	pir|JC1151|JC11	hypothetical 20.3K protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain PO22) plasmid Ti	64	50	354
40	1	3	449	gi|46970	epiD gene product (Staphylococcus epidermidis)	64	41	447
40	7	4683	4976	gnl|PID|e325792	(AJ000005) glucose kinase (Bacillus megaterium)	64	45	294
45	7	8068	6920	gnl|PID|d102036	subunit of ADP-glucose pyrophosphorylase (Bacillus stearothermophilus)	64	40	1149
51	2	301	1059	gi|43985	nifS-like gene (Lactobacillus delbrueckii)	64	54	759
51	13	15251	18397	gi|2293260	(AF008220) DNA-polymerase III alpha-chain (Bacillus subtilis)	64	46	3147
53	3	1157	555	gi|1574292	hypothetical (Haemophilus influenzae)	64	47	603
58	2	4236	1606	gi|1573826	alanyl-tRNA synthetase (alaS) (Haemophilus influenzae)	64	51	2631
66	1	3	1259	gi|895749	putative cellobiose phosphotransferase enzyme II'' (Bacillus subtilis)	64	42	1257
68	5	5213	6556	gi|436965	(malA) gene products (Bacillus stearothermophilus)	64	47	1344
69	6	5356	4949	gnl|PID|d101316	Cdd (Bacillus subtilis)	64	52	408

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
74	4	6948	5038	gi|726480	L-glutamine-D-fructose-6-phosphate amidotransferase (Bacillus subtilis)	64	50	1911
75	3	1283	1465	bbs|133379	TLS-CHOP=fusion protein(CHOP=C/EBP transcription factor, TLS=nuclear RNA-binding protein) (human, myxoid liposarcomas cells, Peptide Mutant, 462 aa) (Homo sapiens)	64	57	183
81	13	14016	14231	gi|143175	methanol dehydrogenase alpha-10 subunit (Bacillus sp.)	64	35	216
83	22	21851	22090	gnl|PID|d101315	YqfA (Bacillus subtilis)	64	44	240
87	11	10046	9300	gnl|PID|e323505	putative PtcI protein (Bacillus subtilis)	64	43	747
98	7	5032	5706	gnl|PID|e233880	hypothetical protein (Bacillus subtilis)	64	38	675
105	1	2	1276	gi|1657503	similar to S. aureus mercury(II) reductase (Escherichia coli)	64	45	1275
111	7	5136	6410	gnl|PID|d101119	NifS (Synechocystis sp.)	64	50	1275
119	1	2	1297	gnl|PID|e320520	hypothetical protein (Natronobacterium pharaonis)	64	37	1296
123	3	1125	2156	gnl|PID|e253284	ORF YDL244w (Saccharomyces cerevisiae)	64	40	1032
124	5	2331	1780	gnl|PID|d101884	hypothetical protein (Synechocystis sp.)	64	50	552
129	4	3467	2709	gnl|PID|d101314	YqeU (Bacillus subtilis)	64	52	759
131	1	152	3	gi|1377841	unknown (Bacillus subtilis)	64	42	150
137	11	7196	7549	pir|JC1151|JC11	hypothetical 20.3K protein (insertion sequence ISI131) - Agrobacterium tumefaciens (strain PO22) plasmid Ti	64	50	354
139	3	3226	2651	gi|2293301	(AF008220) YtqB (Bacillus subtilis)	64	44	576
146	10	6730	5648	gi|1322245	mevalonate pyrophosphate decarboxylase (Rattus norvegicus)	64	45	1083
147	1	2	1018	gnl|PID|e137033	unknown gene product (Lactobacillus leichmannii)	64	46	1017
148	11	8430	8783	gi|2130630	(AF000430) dynamin-like protein (Homo sapiens)	64	28	354
156	7	4313	3612	gnl|PID|d102050	transmembrane (Bacillus subtilis)	64	31	702
157	4	1299	2114	gnl|PID|d100892	homologous to Gln transport system permease proteins (Bacillus subtilis)	64	43	816
162	6	5880	6362	gi|517204	ORF1, putative 42 kDa protein (Streptococcus pyogenes)	64	58	483
164	13	9707	8769	gnl|PID|d100964	homologue of ferric anguibactin transport system permerase protein FatD of V. anguillarum (Bacillus subtilis)	64	40	939
175	5	3906	4598	gi|534045	antiterminator (Bacillus subtilis)	64	39	693
189	10	6154	6507	gi|581307	response regulator (Lactobacillus plantarum)	64	33	354
191	4	3519	2863	gi|149520	phosphoribosyl anthranilate isomerase (Lactococcus lactis)	64	46	657

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins Similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
202	1	76	1140	|gnl|PID|e293806	O-acetylhomoserine sulfhydrylase (Leptospira meyeri)	64	47	1065
224	1	234	1571	|gi|1573393	collagenase (prtC) (Haemophilus influenzae)	64	42	1338
231	3	291	647	|gi|40174	ORF X (Bacillus subtilis)	64	43	357
253	3	709	1089	|pir|JC1151|JC11	hypothetical 20.3K protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain PO22) plasmid Ti	64	50	381
265	1	820	2	|gi|1377832	unknown (Bacillus subtilis)	64	31	819
297	1	1	660	|gi|1590871	collagenase (Methanococcus jannaschii)	64	48	660
328	1	263	21	|gi|992651	Gln4p (Saccharomyces cerevisiae)	64	41	243
5	4	8730	8098	|gi|556885	Unknown (Bacillus subtilis)	63	48	633
10	6	5178	4483	|gi|1573101	hypothetical (Haemophilus influenzae)	63	40	696
12	11	9324	9902	|gi|806536	membrane protein (Bacillus acidopullulyticus)	63	42	579
15	10	8897	9187	|gi|722339	unknown (Acetobacter xylinum)	63	40	291
17	2	1031	309	|gnl|PID|e217602	PlnU (Lactobacillus plantarum)	63	32	723
18	8	7778	6975	|gi|1377843	unknown (Bacillus subtilis)	63	45	804
26	4	9780	7078	|gi|142440	ATP-dependent nuclease (Bacillus subtilis)	63	46	2703
29	5	3488	4192	|gi|1377829	unknown (Bacillus subtilis)	63	35	705
34	11	8830	7988	|gnl|PID|d101198	ORF8 (Enterococcus faecalis)	63	45	843
35	3	1187	876	|gi|722339	unknown (Acetobacter xylinum)	63	39	312
48	15	12509	11691	|gi|1573389	hypothetical (Haemophilus influenzae)	63	41	819
51	11	12719	12189	|gi|142450	ahrC protein (Bacillus subtilis)	63	35	531
55	4	3979	5022	|gi|1708640	YeaB (Bacillus subtilis)	63	41	1044
55	15	13669	14670	|gnl|PID|e311502	thioredoxine reductase (Bacillus subtilis)	63	44	1002
68	10	9242	8919	|sp|P37686|YIAY_	HYPOTHETICAL 40.2 KD PROTEIN IN AVTA-SELB INTERGENIC REGION (F382).	63	40	324
86	7	6554	5685	|gi|1574382	lic-1 operon protein (licD) (Haemophilus influenzae)	63	41	870
88	8	6085	5180	|gi|2098719	putative fimbrial-associated protein (Actinomyces naeslundii)	63	43	906
96	8	5858	6484	|gi|1052803	orf1gyrb gene product (Streptococcus pneumoniae)	63	38	627
100	1	240	1940	|gi|7171	fucosidase (Dictyostelium discoideum)	63	36	1701

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
104	4	3063	5765	gi|144985	phosphoenolpyruvate carboxylase (Corynebacterium glutamicum)	63	46	2703
106	8	9189	8554	gi|533099	endonuclease III (Bacillus subtilis)	63	45	636
122	6	4704	4886	gnl|PID|d101139	transposase (Synechocystis sp.)	63	39	183
128	7	4517	5203	gnl|PID|d101414	orf2 (Methanobacterium thermoautotrophicum)	63	50	687
137	4	963	1547	gi|472920	v-type Na-ATPase (Enterococcus hirae)	63	27	585
142	7	4100	4585	gnl|PID|e313025	hypothetical protein (Bacillus subtilis)	63	44	486
159	5	1741	2571	gi|1787043	(AE000184) f271; This 271 aa orf is 24 pct identical (16 gaps) to 265 residues of an approx. 272 aa protein YIDA_ECOLI SW: P09997 (Escherichia coli)	63	39	831
171	12	8803	14406	gnl|PID|e324918	IgA1 protease (Streptococcus sanguis)	63	48	5604
177	1	3	347	gi|1773150	hypothetical 14.8kd protein (Escherichia coli)	63	34	345
178	2	423	917	gi|722339	unknown (Acetobacter xylinum)	63	41	495
178	3	794	1012	gi|1591582	cobalamin biosynthesis protein N (Methanococcus jannaschii)	63	36	219
195	1	1377	175	gnl|PID|e324217	ftsQ (Enterococcus hirae)	63	33	1203
234	5	1739	1527	gi|1591582	cobalamin biosynthesis protein N (Methanococcus jannaschii)	63	36	213
249	1	81	257	gi|1000453	TreR (Bacillus subtilis)	63	41	177
283	1	127	1347	gi|396486	ORF8 (Bacillus subtilis)	63	44	1221
293	3	2804	3466	gi|722339	unknown (Acetobacter xylinum)	63	37	663
311	1	905	486	gi|1877424	UDP-galactose 4-epimerase (Streptococcus mutans)	63	46	420
324	1	2	556	gi|1477741	histidine periplasmic binding protein P29 (Campylobacter jejuni)	63	36	555
365	1	219	13	gi|2252843	(AF013293) No definition line found (Arabidopsis thaliana)	63	33	207
382	1	88	378	gi|722339	unknown (Acetobacter xylinum)	63	40	291
385	3	364	158	gi|2252843	(AF013293) No definition line found (Arabidopsis thaliana)	63	33	207
2	1	2495	288	gnl|PID|e325007	penicillin-binding protein (Bacillus subtilis)	62	42	2208
3	23	23374	24231	gnl|PID|e254993	hypothetical protein (Bacillus subtilis)	62	35	858
6	16	14320	13193	gnl|PID|e349614	nifS-like protein (Mycobacterium leprae)	62	37	1128
7	8	6819	7232	gnl|PID|d101324	YqhY (Bacillus subtilis)	62	32	414
7	19	15466	14207	gnl|PID|d101804	beta ketoacyl-acyl carrier protein synthase (Synechocystis sp.)	62	43	1260

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
7	21	17155	16229	gnl|PID|e323514	putative FabD protein (Bacillus subtilis)	62	46	927
7	24	19526	18519	gi|1276434	beta-ketoacyl-ACP synthase III (Cuphea wrightii)	62	37	1008
12	7	5904	4702	gi|1573768	A/G-specific adenine glycosylase (mutY) (Haemophilus influenzae)	62	43	1203
12	9	8032	8793	gi|1591587	pantothenate metabolism flavoprotein (Methanococcus jannaschii)	62	33	762
15	11	9678	9328	pir|JC1151|JC11	hypothetical 20.3K protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain PO22) plasmid Ti	62	43	351
17	4	2609	2442	gi|1591081	M. jannaschii predicted coding region MJ0374 (Methanococcus jannaschii)	62	43	168
17	5	3053	2835	gi|149570	role in the expression of lactacin F, part of the laf operon (Lactobacillus sp.)	62	44	219
22	10	8627	9538	gnl|PID|d100580	similar to B. subtilis DnaH (Bacillus subtilis)	62	43	912
30	3	865	2043	gi|2314379	(AE000627) ABC transporter, ATP-binding protein (yhcG) (Helicobacter pylori)	62	43	1179
33	5	2235	1636	gi|413976	ipa-52r gene product (Bacillus subtilis)	62	44	600
38	11	5689	6123	gi|148231	o251 (Escherichia coli)	62	34	435
40	17	14272	13328	gnl|PID|d101904	hypothetical protein (Synechocystis sp.)	62	43	945
42	1	3	311	gi|1146182	putative (Bacillus subtilis)	62	41	309
44	2	1267	4005	gi|1786952	(AE000176) o877; 100 pct identical to the first 86 residues of the 100 aa hypothetical protein fragment YBGB_ECOLI SW: P54746 (Escherichia coli)	62	43	2739
48	12	9732	9304	gi|662920	repressor protein (Enterococcus hirae)	62	32	429
51	8	5664	7181	gnl|PID|e301153	StySKI methylase (Salmonella enterica)	62	44	1518
52	3	2791	2099	gi|1183886	integral membrane protein (Bacillus subtilis)	62	41	693
55	16	15702	14704	gnl|PID|e313028	hypothetical protein (Bacillus subtilis)	62	40	999
59	6	3418	3984	gi|2065483	unknown (Lactococcus lactis lactis)	62	32	567
63	5	4997	4809	gi|149771	pilin gene inverting protein (PivML) (Moraxella lacunata)	62	28	189
70	14	10002	10739	gi|992977	bplG gene product (Bordetella pertussis)	62	45	738
71	13	18790	20382	gi|1280135	coded for by C. elegans cDNA cm21e6; coded for by C. elegans cDNA cm01e2; similar to melibiose carrier protein (thiomethylgalactoside permease II) (Caenorhabditis elegans)	62	62	1593
71	28	32217	32768	gnl|PID|d101312	YqeG (Bacillus subtilis)	62	35	552
74	7	11666	10383	gi|1552753	hypothetical (Escherichia coli)	62	38	1284

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
80	8	9370	9609	|gnl|PID|d102002	(AB001488) FUNCTION UNKNOWN. [Bacillus subtilis]	62	46	240
97	10	9068	7041	|gi|88246)	protein-N(pi)-phosphohistidine-sugar phosphotransferase [Escherichia coli]	62	42	2028
98	4	2306	3268	|gnl|PID|d101496	BraE (integral membrane protein) [Pseudomonas aeruginosa]	62	42	963
102	3	2823	3539	|gnl|PID|e313010	hypothetical protein [Bacillus subtilis]	62	24	717
103	3	2795	1242	|gnl|PID|d102049	H. influenzae hypothetical ABC transporter; P44808 (974) [Bacillus subtilis]	62	41	1554
111	2	2035	3462	|gi|581297	NisP [Lactococcus lactis]	62	44	1428
112	4	3154	4080	|gi|1574379	lic-1 operon protein (licA) [Haemophilus influenzae]	62	39	927
112	6	4939	5649	|gi|1574381	lic-1 operon protein (licC) [Haemophilus influenzae]	62	39	711
124	3	1137	721	|gi|1573024	anaerobic ribonucleoside-triphosphate reductase (nrdD) [Haemophilus influenzae]	62	45	417
124	6	3162	2329	|gi|609076	leucyl aminopeptidase [Lactobacillus delbrueckii]	62	40	834
126	7	11073	7516	|gnl|PID|d101163	ORF4 [Bacillus subtilis]	62	38	3558
129	6	4983	4540	|pir|S41509|S415	zinc finger protein EF6 - Chilo iridescent virus	62	48	444
131	7	4510	4103	|gi|1857245	unknown [Lactococcus lactis]	62	42	408
149	2	1923	2579	|gi|1592142	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	62	41	657
149	7	5360	6055	|gnl|PID|e323508	YloS protein [Bacillus subtilis]	62	40	696
156	1	450	238	|gnl|PID|e254644	membrane protein [Streptococcus pneumoniae]	62	40	213
156	6	3606	2935	|gnl|PID|d102050	transmembrane [Bacillus subtilis]	62	37	672
171	2	1779	2291	|gi|43941	EIII-B Sor PTS [Klebsiella pneumoniae]	62	35	513
172	2	385	723	|gi|895750	putative cellobiose phosphotransferase enzyme III [Bacillus subtilis]	62	39	339
173	3	2599	893	|gi|1591732	cobalt transport ATP-binding protein O [Methanococcus jannaschii]	62	42	1707
179	2	492	1754	|gi|1574071	H. influenzae predicted coding region HI1038 [Haemophilus influenzae]	62	38	1263
181	6	2856	3707	|gi|1777435	LacT [Lactobacillus casei]	62	42	852
185	2	2074	311	|gi|2182397	(AE000073) Y4fN [Rhizobium sp. NGR234]	62	41	1764
200	2	1061	1984	|gi|450566	transmembrane protein [Bacillus subtilis]	62	37	924
202	3	2583	3473	|gi|42219	P35 gene product (AA 1 - 314) [Escherichia coli]	62	41	891
210	3	1374	1565	|gi|49315	ORF1 gene product [Bacillus subtilis]	62	45	192

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
211	1	3	971	gi|147402	mannose permease subunit III-Man (Escherichia coli)	62	43	969
223	2	1495	1034	gnl|PID|d101190	ORF2 (Streptococcus mutans)	62	41	462
228	1	34	909	gi|530063	glycerol uptake facilitator (Streptococcus pneumoniae)	62	44	876
234	2	90	917	gi|2293259	(AF008220) YtqI (Bacillus subtilis)	62	38	828
282	5	1765	1487	gnl|PID|e276475	galactokinase (Arabidopsis thaliana)	62	33	279
375	1	1	159	gi|1674211	(AE000052) Mycoplasma pneumoniae, hypothetical protein homolog; similar to Swiss-Prot Accession Number P35155, from B. subtilis (Mycoplasma pneumoniae)	62	40	159
385	5	584	357	gi|1573353	outer membrane integrity protein (tolA) (Haemophilus influenzae)	62	47	228
3	19	18550	19269	gi|606162	ORF_f229 (Escherichia coli)	61	41	720
7	4	2725	3225	gi|2114425	similar to Synechocystis sp. hypothetical protein, encoded by GenBank Accession Number D64006 (Bacillus subtilis)	61	42	501
17	6	3326	3054	gi|149569	lactacin F (Lactobacillus sp.)	61	43	273
44	3	4061	4957	gnl|PID|d101068	xylose repressor (Synechocystis sp.)	61	38	897
54	11	8388	7234	gnl|PID|d101329	YqjH (Bacillus subtilis)	61	42	1155
57	6	3974	6037	gnl|PID|d101316	YqfK (Bacillus subtilis)	61	42	2064
58	5	7356	6565	sp|P45169|POTC_	SPERMIDINE/PUTRESCINE TRANSPORT SYSTEM PERMEASE PROTEIN POTC.	61	34	792
67	1	3	692	gi|537108	ORF_f254 (Escherichia coli)	61	46	690
68	9	8816	7890	gi|19501	pPLZ12 gene product (AA 1-184) (Lupinus polyphyllus)	61	41	927
70	15	10737	12008	gi|992976	bplF gene product (Bordetella pertussis)	61	44	1272
72	11	9759	10202	gnl|PID|d101833	carboxynorspermidine decarboxylase (Synechocystis sp.)	61	36	444
76	8	7881	7003	gnl|PID|d100305	farnesyl diphosphate synthase (Bacillus stearothermophilus)	61	45	879
87	4	4914	3697	gi|528991	unknown (Bacillus subtilis)	61	42	1218
87	13	12311	11361	gi|1789683	(AE000407) methionyl-tRNA formyltransferase (Escherichia coli)	61	44	951
91	2	711	2989	gi|537080	ribonucleoside triphosphate reductase (Escherichia coli)	61	45	2259
105	3	2711	3499	gnl|PID|d101851	hypothetical protein (Synechocystis sp.)	61	44	789
115	6	7968	6478	gi|895747	putative cel operon regulator (Bacillus subtilis)	61	36	1491
123	8	7181	8518	gi|1209527	protein histidine kinase (Enterococcus faecalis)	61	40	1338

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
126	6	7525	6725	gi|1787043	(AE000184) f271; This 271 aa orf is 24 pct identical (16 gaps) to 265 residues of an approx. 272 aa protein YIDA_ECOLI SW: P09997 (Escherichia coli)	61	38	801
128	1	1	639	gnl|PID|d101328	YqiY (Bacillus subtilis)	61	41	639
139	7	4794	5054	gi|1022726	unknown (Staphylococcus haemolyticus)	61	41	261
139	9	12632	5913	gnl|PID|e270014	beta-galactosidase (Thermoanaerobacter ethanolicus)	61	41	6720
141	1	2552	42	gi|520541	penicillin-binding proteins 1A and 1B (Bacillus subtilis)	61	42	2511
148	16	12125	11424	gi|1552743	tetrahydrodipicolinate N-succinyltransferase (Escherichia coli)	61	42	702
162	3	4112	3456	gnl|PID|d101829	phosphoglycolate phosphatase (Synechocystis sp.)	61	30	657
172	3	727	1077	gnl|PID|d102048	B. subtilis, cellobiose phosphotransferase system, celA; P46318 (220) (Bacillus subtilis)	61	44	351
177	3	1101	1772	gnl|PID|d100574	unknown (Bacillus subtilis)	61	43	672
202	2	1278	2585	gi|1045831	hypothetical protein (GB:L18965_6) (Mycoplasma genitalium)	61	36	1308
224	3	2782	3144	gi|1591144	M. jannaschii predicted coding region MJ0440 (Methanococcus jannaschii)	61	30	363
225	4	3395	3766	gi|1552774	hypothetical (Escherichia coli)	61	40	372
249	2	212	802	gi|1000453	TreR (Bacillus subtilis)	61	42	591
254	2	843	484	gnl|PID|d100417	ORF120 (Escherichia coli)	61	36	360
257	1	3	350	gnl|PID|e255315	unknown (Mycobacterium tuberculosis)	61	42	348
293	4	3971	3657	pir|JC1151|JC11	hypothetical 20.3K protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain P022) plasmid Ti	61	45	315
301	1	949	17	gi|2291209	(AF016424) contains similarity to acyltransferases (Caenorhabditis elegans)	61	33	933
373	1	1066	287	gi|393396	Tb-292 membrane associated protein (Trypanosoma brucei subgroup)	61	38	780
3	24	24473	24955	gi|537093	ORF_o153b (Escherichia coli)	60	27	483
6	5	4636	5739	gi|2293258	(AF008220) YtoI (Bacillus subtilis)	60	35	1104
6	12	11936	11187	gi|293017	ORF3 (put.); putative (Lactococcus lactis)	60	44	750
17	13	6708	6484	gi|149569	lactacin F (Lactobacillus sp.)	60	32	225
18	7	6977	5670	gi|1788140	(AE000278) o481; This 481 aa orf is 35 pct identical (19 gaps) to 309 residues of an approx. 856 aa protein NOL1_HUMAN SW: P46087 (Escherichia coli)	60	43	1308
20	15	15878	17167	gnl|PID|d100584	unknown (Bacillus subtilis)	60	44	1290

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
22	1	1	243	gnl|PID|d102050	transmembrane [Bacillus subtilis]	60	36	243
32	10	8296	8964	gi|2293275	(AF008220) YtaG [Bacillus subtilis]	60	37	669
38	15	8837	9697	gi|40023	B.subtilis genes rpmH, rnpA, 50kd, gidA and gidB [Bacillus subtilis]	60	35	861
43	6	8610	5944	gi|171787	protein kinase 1 [Saccharomyces cerevisiae]	60	36	2667
44	1	1	1269	gnl|PID|e235823	unknown [Schizosaccharomyces pombe]	60	44	1269
45	10	11138	10368	gi|397488	1,4-alpha-glucan branching enzyme [Bacillus subtilis]	60	43	771
48	19	15766	14378	gnl|PID|e205173	orf1 [Lactobacillus helveticus]	60	39	1389
48	21	16727	16951	gnl|PID|d102041	(AB002668) unnamed protein product [Haemophilus actinomycetemcomitans]	60	32	225
50	1	2	898	gnl|PID|e246537	ORF286 protein [Pseudomonas stutzeri]	60	31	897
62	2	638	1177	gnl|PID|d100587	unknown [Bacillus subtilis]	60	42	540
68	4	3590	5203	gi|1573583	H. influenzae predicted coding region HI0594 [Haemophilus influenzae]	60	36	1614
70	11	5781	6182	gnl|PID|d102014	(AB001488) SIMILAR TO YDFR GENE PRODUCT OF THIS ENTRY (YDFR_BACSU). [Bacillus subtilis]	60	33	402
70	12	6343	8133	gnl|PID|e324970	hypothetical protein [Bacillus subtilis]	60	38	1791
71	8	11701	14157	gi|580866	ipa-12d gene product [Bacillus subtilis]	60	33	2457
74	8	12509	11664	gnl|PID|d101832	phosphatidate cytidylyltransferase [Synechocystis sp.]	60	45	846
76	4	4116	3367	gi|2352096	orf; similar to serine/threonine protein phosphatase [Fervidobacterium islandicum]	60	39	750
80	4	7372	7665	gi|1786420	(AE000131) f86; 100 pct identical to GB: ECODINJ_6 ACCESSION: D38582 [Escherichia coli]	60	30	294
81	6	4073	4522	gi|147402	mannose permease subunit III-Man [Escherichia coli]	60	35	450
86	1	940	155	gi|143177	putative [Bacillus subtilis]	60	26	786
92	1	1	192	gi|396348	homoserine transsuccinylase [Escherichia coli]	60	45	192
93	14	10619	9384	gi|1788389	(AE000297) o464; This 464 aa orf is 33 pct identical (9 gaps) to 331 residues of an approx. 416 aa protein HTRC_NEIGO SW: P43505 [Escherichia coli]	60	27	1236
94	5	5548	8121	gnl|PID|e329895	(AJ000496) cyclic nucleotide-gated channel beta subunit [Rattus norvegicus]	60	50	2574
97	7	5396	4533	gi|1591396	transketolase' [Methanococcus jannaschii]	60	43	864
102	2	2081	2833	gnl|PID|e320929	hypothetical protein [Mycobacterium tuberculosis]	60	43	753

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
106	9	9773	9183	|gnl|PID|e334782	YlbN protein [Bacillus subtilis]	60	31	591
113	8	6361	6837	|gi|466875	nifU; B1496_C1_157 [Mycobacterium leprae]	60	43	477
115	2	2755	524	|gnl|PID|e328143	(AJ000332) Glucosidase II [Homo sapiens]	60	32	2232
122	7	4763	5068	|gnl|PID|d101876	transposase [Synechocystis sp.]	60	39	306
127	8	4510	5283	|gi|1777938	Pgm [Treponema pallidum]	60	38	774
138	4	3082	2672	|gnl|PID|e325196	hypothetical protein [Bacillus subtilis]	60	36	411
139	1	177	4	|gnl|PID|d100680	ORF [Thermus thermophilus]	60	39	174
139	11	14520	13009	|gi|537145	ORF_f437 [Escherichia coli]	60	30	1512
140	2	2592	1249	|gi|1209527	protein histidine kinase [Enterococcus faecalis]	60	37	1344
141	1	210	1049	|gi|463181	E5 ORF from bp 3842 to 4081; putative [Human papillomavirus type 33]	60	34	840
141	5	5368	6405	|gi|145362	tyrosine-sensitive DAHP synthase (aroF) [Escherichia coli]	60	41	1038
142	6	3558	4049	|gi|600711	putative [Bacillus subtilis]	60	37	492
148	10	7742	8713	|gnl|PID|e313022	hypothetical protein [Bacillus subtilis]	60	27	972
153	5	3667	4278	|gi|2293322	(AF008220) branch-chain amino acid transporter [Bacillus subtilis]	60	42	612
155	1	1413	748	|gi|2104504	putative UDP-glucose dehydrogenase [Escherichia coli]	60	40	666
158	3	3116	2472	|gnl|PID|d100872	a negative regulator of pho regulon [Pseudomonas aeruginosa]	60	37	645
159	3	778	1386	|gnl|PID|e308090	product highly similar to Bacillus anthracis CapA protein [Bacillus subtilis]	60	48	609
163	7	8049	8468	|gnl|PID|d101313	YqeN [Bacillus subtilis]	60	38	420
170	3	4130	2688	|gi|1574179	H. influenzae predicted coding region HI1244 [Haemophilus influenzae]	60	39	1443
171	7	4717	5901	|gi|606076	ORF_o384 [Escherichia coli]	60	44	1185
183	3	2440	2135	|gi|1877427	repressor [Streptococcus pyogenes phage T12]	60	38	306
191	10	9444	8428	|gi|415664	catabolite control protein [Bacillus megaterium]	60	42	1017
200	1	139	1083	|gi|438462	transmembrane protein [Bacillus subtilis]	60	37	945
201	3	3895	1928	|gi|475112	enzyme IIabc [Pediococcus pentosaceus]	60	39	1968
214	15	10930	10439	|gi|1573407	hypothetical [Haemophilus influenzae]	60	39	492
218	4	2145	2363	|gi|608520	myosin heavy chain kinase A [Dictyostelium discoideum]	60	31	219

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
226	4	2518	2351	gi|437705	hyaluronidase (Streptococcus pneumoniae)	60	53	168
242	1	725	3	gi|43938	Sor regulator (Klebsiella pneumoniae)	60	41	723
245	1	1	288	gi|304897	EcoE type I restriction modification enzyme M subunit (Escherichia coli)	60	56	288
251	1	905	45	gi|671632	unknown (Staphylococcus aureus)	60	36	861
259	1	969	82	gi|153794	rgg (Streptococcus gordonii)	60	32	888
260	2	1492	1662	pir|S31840|S318	probable transposase - Bacillus stearothermophilus	60	26	171
274	1	836	96	gi|1592173	N-ethylammeline chlorohydrolase (Methanococcus jannaschii)	60	40	741
308	1	463	2	gi|1787397	(AE000214) o157 (Escherichia coli)	60	43	462
318	1	3	308	gnl|PID|e137594	xerC recombinase (Lactobacillus leichmannii)	60	42	306
344	1	73	522	gi|509672	repressor protein (Bacteriophage Tuc2009)	60	32	450
5	1	576	4	gi|2293147	(AF008220) YtxM (Bacillus subtilis)	59	31	573
7	22	18140	17142	gnl|PID|e280724	unknown (Mycobacterium tuberculosis)	59	39	999
10	1	1413	4	gi|1353880	sialidase L (Macrobdella decora)	59	41	1410
15	6	6463	5156	gi|580841	F1 (Bacillus subtilis)	59	35	1308
22	2	479	1393	gi|142469	als operom regulatory protein (Bacillus subtilis)	59	34	915
22	5	2698	4614	gnl|PID|e280623	PCPA (Streptococcus pneumoniae)	59	44	1917
30	1	208	558	gnl|PID|e233868	hypothetical protein (Bacillus subtilis)	59	37	351
30	4	3678	2455	gnl|PID|e202290	unknown (Lactobacillus sake)	59	33	1224
35	13	12201	11071	gnl|PID|e238664	hypothetical protein (Bacillus subtilis)	59	35	1131
35	14	13288	12182	gi|1657647	Cap8H (Staphylococcus aureus)	59	39	1107
36	18	18076	17897	gi|1500535	M. jannaschii predicted coding region MJ1635 (Methanococcus jannaschii)	59	33	180
38	12	6172	7137	gi|2293239	(AF008220) YtxK (Bacillus subtilis)	59	34	966
42	3	1952	3361	gi|1684845	pinin (Canis familiaris)	59	40	1410
50	3	2678	1728	gnl|PID|d101329	YqjK (Bacillus subtilis)	59	41	951
56	5	1870	2388	gnl|PID|e137594	xerC recombinase (Lactobacillus leichmannii)	59	41	519
61	6	6812	5620	gnl|PID|e311516	aminotransferase (Bacillus subtilis)	59	40	1185
67	5	2382	3023	gi|1146190	2-keto-3-deoxy-6-phosphogluconate aldolase (Bacillus subtilis)	59	36	642

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
69	10	8567	8899	gi|1573628	antothenate kinase (coaA) (Haemophilus influenzae)	59	38	333
87	12	11383	10055	gnl|PID|e323504	putative Fmu protein (Bacillus subtilis)	59	44	1329
113	14	13927	15894	gi|1673731	(AE000010) Mycoplasma pneumoniae, fructose-permease IIBC component; similar to Swiss-Prot Accession Number P20966, from E. coli (Mycoplasma pneumoniae)	59	43	1968
115	8	8766	8521	gi|1590886	M. jannaschii predicted coding region MJ0110 (Methanococcus jannaschii)	59	38	246
119	2	1966	1526	gnl|PID|e209005	homologous to ORF2 in nrdEF operons of E.coli and S.typhimurium (Lactococcus lactis)	59	43	441
128	17	13438	13178	gnl|PID|e279632	unknown (Mycobacterium tuberculosis)	59	38	261
140	22	23903	23388	gi|482922	protein with homology to pail repressor of B.subtilis (Lactobacillus delbrueckii)	59	40	516
148	13	9697	9014	gnl|PID|d102005	(AB001488) FUNCTION UNKNOWN, SIMILAR PRODUCT IN H. INFLUENZAE AND SYNECHOCYSTIS. (Bacillus subtilis)	59	32	684
149	10	7213	8244	gi|710422	cmp-binding-factor 1 (Staphylococcus aureus)	59	40	1032
164	9	6993	6013	gnl|PID|d100965	ferric anguibactin-binding protein precusor FatB of V. anguillarum (Bacillus subtilis)	59	41	981
164	12	8836	7823	gnl|PID|d100964	homologue of ferric anguibactin transport system permerase protein FatC of V. anguillarum (Bacillus subtilis)	59	35	1014
177	2	401	1072	gi|289759	coded for by C. elegans cDNA CE2G3 (GenBank:Z14728); putative (Caenorhabditis elegans)	59	40	672
177	7	3841	4200	gi|2313445	(AE000551) H pylori predicted coding region HP0342 (Helicobacter pylori)	59	38	360
183	4	2768	2508	gi|509672	repressor protein (Bacteriophage Tuc2009)	59	50	261
186	6	3398	2820	gi|606080	ORF_o290; Geneplot suggests frameshift linking to o267, not found (Escherichia coli)	59	38	579
190	3	3120	1711	gi|1613768	histidine protein kinase (Streptococcus pneumoniae)	59	32	1410
194	2	1621	1019	gnl|PID|d100579	unknown (Bacillus subtilis)	59	40	603
198	7	5205	4306	gnl|PID|e313073	hypothetical protein (Bacillus subtilis)	59	38	900
220	5	4362	3958	gnl|PID|d101322	YqhL (Bacillus subtilis)	59	46	405
242	3	1573	2367	gi|1787045	(AE000184) f308; This 308 aa orf is 35 pct identical (35 gaps) to 305 residues of an approx. 296 aa protein PFLC_ECOLI SW: P32675 (Escherichia coli)	59	42	795
247	2	1154	1480	gi|40073	ORF107 (Bacillus subtilis)	59	39	327

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
256	1	868	2	|gnl|PID|d101924	|hemolysin (Synechocystis sp.)	59	39	867
258	1	65	820	|gi|2246532	|ORF 73, contains large complex repeat CR 73 (Kaposi's sarcoma-associated herpesvirus)	59	20	756
270	1	386	1126	|gnl|PID|d102092	|YfnB (Bacillus subtilis)	59	40	741
281	1	552	166	|gi|666062	|putative (Lactococcus lactis)	59	31	387
309	1	3	479	|gi|405879	|yeiH (Escherichia coli)	59	38	477
363	1	2	1894	|gi|915208	|gastric mucin (Sus scrofa)	59	31	1893
387	2	425	84	|gi|160671	|S antigen precursor (Plasmodium falciparum)	59	44	342
5	6	11223	10465	|gnl|PID|d101812	|LumQ (Synechocystis sp.)	58	29	759
29	4	2098	3513	|gnl|PID|d100479	|Na+ -ATPase subunit J (Enterococcus hirae)	58	39	1416
30	5	4058	3651	|gi|39478	|ATP binding protein of transport ATPases (Bacillus firmus)	58	34	408
33	6	2983	2210	|gnl|PID|d101164	|unknown (Bacillus subtilis)	58	45	774
36	8	5316	6179	|gi|1518679	|orf (Bacillus subtilis)	58	32	864
43	5	5926	3971	|gi|1788150	|(AE000278) protease II (Escherichia coli)	58	37	1956
46	5	3704	5221	|gnl|PID|e267329	|Unknown (Bacillus subtilis)	58	42	1518
48	14	11722	11066	|gnl|PID|d101771	|thiamin biosynthetic bifunctional enzyme (Synechocystis sp.)	58	34	657
52	1	1229	3	|gnl|PID|d101291	|reductase (Pseudomonas aeruginosa)	58	35	1227
53	2	702	412	|gi|2313357	|(AE000545) cytochrome c biogenesis protein (ccdA) (Helicobacter pylori)	58	25	291
58	4	6586	5498	|gi|147329	|transport protein (Escherichia coli)	58	41	1089
69	5	4934	3807	|gnl|PID|e311492	|unknown (Bacillus subtilis)	58	41	1128
71	27	31357	32277	|gi|2408014	|hypothetical protein (Schizosaccharomyces pombe)	58	33	921
72	4	3586	2882	|gi|18694	|nodulin-21 (AA 1-201; (Glycine max)	58	34	705
74	3	4917	4230	|gi|2293252	|(AF008220) YtmO (Bacillus subtilis)	58	33	708
79	4	4594	3422	|gi|1217989	|ORF3 (Streptococcus pneumoniae)	58	44	1173
82	8	10585	8171	|gi|882711	|exonuclease V alpha-subunit (Escherichia coli)	58	38	2415
86	17	16017	15337	|gi|47642	|5-dehydroquinate hydrolyase (3-dehydroquinase) (Salmonella typhi)	58	32	681
97	2	931	560	|gi|153794	|rgg (Streptococcus gordonii)	58	32	372

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
108	2	358	2724	gi|537020	vacB gene product (Escherichia coli)	58	37	2367
111	5	4593	5240	gi|1592142	ABC transporter, probable ATP-binding subunit (Methanococcus jannaschii)	58	36	648
120	3	4421	5110	gnl|PID|d101320	YqgX (Bacillus subtilis)	58	47	690
1	16	13131	12673	gi|662919	ORF U (Enterococcus hirae)	58	42	459
132	3	6174	4939	gi|1800301	macrolide-efflux determinant (Streptococcus pneumoniae)	58	35	1236
133	1	111	890	gnl|PID|e269488	Unknown (Bacillus subtilis)	58	36	780
160	11	8615	9865	gi|473901	ORF1 (Lactococcus lactis)	58	39	1251
161	6	6268	6849	gnl|PID|d101024	DJ-1 protein (Homo sapiens)	58	32	582
169	1	214	2	gnl|PID|d100447	translation elongation factor-1 (Chlorella virus)	58	31	213
187	1	487	2	gi|475114	regulatory protein (Pediococcus pentosaceus)	58	38	486
187	6	4384	4620	gi|167475	dessication-related protein (Craterostigma plantagineum)	58	55	237
190	2	1464	1640	gnl|PID|e246727	competence pheromone (Streptococcus gordonii)	58	38	177
192	2	2012	1344	gnl|PID|d100556	rat GCP360 (Rattus rattus)	58	44	669
206	1	1292	696	gnl|PID|e202579	product similar to WrbA (Lactobacillus sake)	58	35	597
216	2	2333	555	gnl|PID|e325036	hypothetical protein (Bacillus subtilis)	58	33	1779
217	5	5250	4321	gi|466474	cellobiose phosphotransferase enzyme II'' (Bacillus stearothermophilus)	58	38	930
217	7	5636	5106	gnl|PID|d102048	B. subtilis cellobiose phosphotransferase system celB; P46317 (998) transmembrane (Bacillus subtilis)	58	44	531
232	1	2	811	gi|1573777	cell division ATP-binding protein (ftsE) (Haemophilus influenzae)	58	39	810
264	1	2	715	gi|971330	NatA (Bacillus subtilis)	58	32	714
280	1	33	767	gi|1786187	(AE000111) hypothetical 29.6 kD protein in thrC-talB intergenic region (Escherichia coli)	58	31	735
306	1	845	3	gnl|PID|e334780	YIbL protein (Bacillus subtilis)	58	47	843
360	3	1556	1092	sp|P46351|YZGD_	HYPOTHETICAL 45.4 KD PROTEIN IN THIAMINASE I 5'REGION	58	32	465
363	5	2160	1867	gi|160671	S antigen precursor (Plasmodium falciparum)	58	51	294
372	1	806	3	gi|393394	Tb-291 membrane associated protein (Trypanosoma brucei subgroup)	58	37	804
382	2	749	519	pir|JC1151|JC11	hypothetical 20.3K protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain PO22) plasmid Ti	58	41	231

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
3	9	8409	7471	|gi|1499745	(M. jannaschii predicted coding region MJ0912 (Methanococcus jannaschii)	57	38	939
10	10	7674	7507	|gi|1737169	homologue to SKP1 (Arabidopsis thaliana)	57	30	168
11	1	2	412	|gnl|PID|d100139	ORF (Acetobacter pasteurianus)	57	42	411
31	4	2032	1388	|gi|2293213	(AF008220) YtpR (Bacillus subtilis)	57	37	645
33	11	6931	6449	|gnl|PID|e324949	hypothetical protein (Bacillus subtilis)	57	36	483
45	5	5446	5060	|gi|1592204	phosphoserine phosphatase (Methanococcus jannaschii)	57	44	387
49	7	6523	7632	|gi|155369	PTS enzyme-II fructose (Xanthomonas campestris)	57	35	1110
52	6	4520	6850	|gi|1574144	single-stranded-DNA-specific exonuclease (recJ) (Haemophilus influenzae)	57	35	2331
53	5	2079	1795	|gi|1843580	replicase-associated polyprotein (oat blue dwarf virus)	57	46	285
61	6	5312	4995	|gi|2182608	(AE000094) Y4rJ (Rhizobium sp. NGR234)	57	39	318
72	15	13883	13059	|gnl|PID|d100892	homologous to SwissProt:YIDA_ECOLI hypothetical protein (Bacillus subtilis)	57	40	825
79	2	2561	1815	|gnl|PID|d100965	homologue of NADPH-flavin oxidoreductase Frp of V. harveyi (Bacillus subtilis)	57	44	747
82	9	9596	9763	|gi|1206045	short region of similarity to glycerophosphoryl diester phosphodiesterases (Caenorhabditis elegans)	57	35	168
86	16	15371	14493	|gi|1787983	(AE000264) o288; 92 pct identical (1 gaps) to 222 residues of fragment YDIB_ECOLI SW: P28244 (223 aa) (Escherichia coli)	57	34	879
93	3	1695	1177	|gi|1500003	mutator mutT protein (Methanococcus jannaschii)	57	33	519
96	6	3026	4519	|gi|559882	threonine synthase (Arabidopsis thaliana)	57	43	1494
99	14	17211	18212	|gi|773349	BirA protein (Bacillus subtilis)	57	44	1002
112	8	7448	7903	|gi|1591393	(M. jannaschii predicted coding region MJ0678 (Methanococcus jannaschii)	57	30	456
113	16	18627	18328	|pir|A45605|A456	mature-parasite-infected erythrocyte surface antigen MESA - Plasmodium falciparum	57	22	300
123	2	343	1110	|pir|F64149|F641	hypothetical protein HI0355 - Haemophilus influenzae (strain Rd KW20)	57	38	768
123	4	2108	2884	|gnl|PID|d102148	(AB001684) sulfate transport system permease protein (Chlorella vulgaris)	57	39	777
127	10	6477	5587	|gi|1573082	nitrogenase C (nifC) (Haemophilus influenzae)	57	35	891
128	13	9251	9790	|gi|153692	pneumolysin (Streptococcus pneumoniae)	57	38	540
131	4	2139	1363	|gi|42081	nagD gene product (AA 1-250) (Escherichia coli)	57	36	777

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
136	1	214	1221	bbs|148453	SpaA=endocarditis immunodominant antigen [Streptococcus sobrinus, MUCOB 263, Peptide, 1566 aa] [Streptococcus sobrinus]	57	44	1008
140	25	28701	26851	gi|505576	beta-glucoside permease [Bacillus subtilis]	57	38	1851
141	6	6395	7438	gi|995560	unknown [Schizosaccharomyces pombe]	57	41	1044
144	3	3231	2785	gnl|PID|d100139	ORF [Acetobacter pasteurianus]	57	42	447
155	4	5454	4564	gi|600431	glycosyl transerase [Erwinia amylovora]	57	34	891
159	9	4877	5854	gi|290509	o307 [Escherichia coli]	57	35	978
167	11	9710	9249	gnl|PID|d100139	ORF [Acetobacter pasteurianus]	57	42	462
171	6	4023	4436	gi|147402	mannose permease subunit III-Man [Escherichia coli]	57	29	414
178	4	2170	1076	gnl|PID|d102004	(AB001488) ATP-DEPENDENT RNA HELICASE DEAD HOMOLOG. [Bacillus subtilis]	57	39	1095
190	1	145	1455	gi|149420	export/processing protein [Lactococcus lactis]	57	30	1311
198	1	298	95	gi|522268	unidentified ORF22 [Bacteriophage bIL67]	57	36	204
203	2	3195	2110	gnl|PID|e283915	orf c01003 [Sulfolobus solfataricus]	57	41	1086
205	1	40	507	gi|1439527	EIIA-man [Lactobacillus curvatus]	57	28	468
214	7	4243	3797	gnl|PID|d102049	H. influenzae, ribosomal protein alanine acetyltransferase; P44305 (189) [Bacillus subtilis]	57	48	447
268	3	1767	1276	gi|43979	L.curvatus small cryptic plasmid gene for rep protein [Lactobacillus curvatus]	57	36	492
351	1	324	34	gnl|PID|e275871	T03F6.b [Caenorhabditis elegans]	57	31	291
386	1	226	2	gi|160671	S antigen precursor [Plasmodium falciparum]	57	45	225
5	5	10486	8777	gi|405857	yehU [Escherichia coli]	56	33	1710
8	5	3674	3910	gi|467199	pksC; L518_F1_2 [Mycobacterium leprae]	56	39	237
10	3	3442	1874	gnl|PID|d101907	sodium-coupled permease [Synechocystis sp.]	56	36	1569
21	1	1880	333	gi|2313949	(AE000593) osmoprotection protein (proWX) [Helicobacter pylori]	56	33	1548
22	29	21968	22456	gnl|PID|d102001	(AB001488) PROBABLE ACETYLTRANSFERASE. [Bacillus subtilis]	56	37	489
27	1	1361	3	gi|215132	ea59 (525) [Bacteriophage lambda]	56	30	1359
28	9	4667	4278	gi|1592090	DNA repair protein RAD2 [Methanococcus jannaschii]	56	29	390
33	1	3	386	gnl|PID|d100139	ORF [Acetobacter pasteurianus]	56	41	384

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
36	7	5122	5357	pir|PQ0053|PQ00	hypothetical protein (proC 3' region) - Pseudomonas aeruginosa (strain PAO) (fragment)	56	28	276
40	4	3137	4318	gi|1800301	macrolide-efflux determinant (Streptococcus pneumoniae)	56	27	1182
40	16	12511	13191	gnl|PID|e217602	PlnU (Lactobacillus plantarum)	56	38	681
48	17	13775	13023	gi|143729	transcription activator (Bacillus subtilis)	56	35	753
75	4	1674	2594	gnl|PID|d102036	membrane protein (Bacillus stearothermophilus)	56	25	921
85	3	1842	1459	gnl|PID|d100139	ORF (Acetobacter pasteurianus)	56	41	384
89	7	5815	4940	gi|853777	product similar to E coli PRFA2 protein (Bacillus subtilis)	56	42	876
105	2	1360	2718	gnl|PID|d101913	hypothetical protein (Synechocystis sp.)	56	37	1359
112	3	2151	3194	gi|537201	ORF_o345 (Escherichia coli)	56	31	1044
113	4	2754	2963	gnl|PID|d100340	ORF (Plum pox virus)	56	28	210
122	3	1203	2054	gi|1649035	high-affinity periplasmic glutamine binding protein (Salmonella typhimurium)	56	30	852
124	8	3939	3694	gnl|PID|e248893	unknown (Mycobacterium tuberculosis)	56	27	246
125	4	4403	4107	gnl|PID|d100247	human non-muscle myosin heavy chain (Homo sapiens)	56	32	297
127	11	6608	6405	gi|2182397	(AE000073) Y4fN (Rhizobium sp. NGR234)	56	35	204
134	5	4769	3849	gnl|PID|d101870	hypothetical protein (Synechocystis sp.)	56	39	921
137	10	6814	7245	gi|1592011	sulfate permease (cysA) (Methanococcus jannaschii)	56	34	432
142	8	5019	4582	pir|A47071|A470	orf1 immediately 5' of nifS - Bacillus subtilis	56	29	438
146	8	4676	3660	gnl|PID|d101911	hypothetical protein (Synechocystis sp.)	56	32	1017
148	3	1906	2739	gnl|PID|d101099	phosphate transport system permease protein PstA (Synechocystis sp.)	56	36	834
150	4	4449	2743	gnl|PID|e304628	probably site-specific recombinase of the resolvase family of enzymes (Bacteriophage TP21)	56	27	1707
172	1	2	208	gi|1787791	(AE000249) f317; This 317 aa orf is 27 pct identical (16 gaps) to 301 residues of an approx. 320 aa protein YXXC_BACSU SW: P39140 (Escherichia coli)	56	34	207
172	7	4979	5668	gi|396293	similar to Bacillus subtilis hypoth. 20 kDa protein, in tsr 3' region (Escherichia coli)	56	40	690
186	7	3732	3367	gi|1732200	PTS permease for mannose subunit IIPMan (Vibrio furnissii)	56	36	366
187	2	2402	819	pir|S57904|S579	virR49 protein - Streptococcus pyogenes (strain CS101, serotype M49)	56	35	1584

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
204	3	2772	2239	gi|606376	ORF_o162 [Escherichia coli]	56	35	534
206	2	3142	1633	gi|559861	clyM [Plasmid pADI]	56	38	1710
219	3	1689	1096	gi|1146197	putative [Bacillus subtilis]	56	27	594
230	2	409	1485	pir|C60328|C603	hypothetical protein 2 (sr 5' region) - Streptococcus mutans (strain OMZ175, serotype f)	56	40	1077
233	4	2930	3268	gi|1041785	rhoptry protein [Plasmodium yoelii]	56	24	339
273	2	1543	2724	gi|143089	iep protein [Bacillus subtilis]	56	32	1182
353	1	1	516	gnl|PID|e325000	hypothetical protein [Bacillus subtilis]	56	41	516
359	1	87	641	gi|1786952	(AE000176) o877; 100 pct identical to the first 86 residues of the 100 aa hypothetical protein fragment YBGB_ECOLI SW: P54746 [Escherichia coli]	56	46	555
363	7	4482	4198	gi|1573353	outer membrane integrity protein (tolA) [Haemophilus influenzae]	56	38	285
376	1	2	508	gnl|PID|e325031	hypothetical protein [Bacillus subtilis]	56	33	507
18	1	836	177	gnl|PID|d100872	a negative regulator of pho regulon [Pseudomonas aeruginosa]	55	31	660
28	4	1824	1618	gnl|PID|e316518	STAT protein [Dictyostelium discoideum]	55	40	207
29	6	4496	5041	gi|1088261	unknown protein [Anabaena sp.]	55	31	546
38	16	9695	10702	gi|580905	B.subtilis genes rpmH, rnpA, 50kd, gidA and gidB [Bacillus subtilis]	55	31	1008
49	5	5727	6182	gi|1786951	(AE000176) heat-responsive regulatory protein [Escherichia coli]	55	29	456
51	4	2381	3241	gnl|PID|d101293	YbbA [Bacillus subtilis]	55	42	861
52	9	9640	10866	gi|153016	ORF 419 protein [Staphylococcus aureus]	55	23	1227
53	4	1813	1349	gi|896042	OspF [Borrelia burgdorferi]	55	30	465
60	5	4794	5756	gi|1499876	magnesium and cobalt transport protein [Methanococcus jannaschii]	55	38	963
71	9	14176	15408	gi|1857120	glycosyl transferase [Neisseria meningitidis]	55	41	1233
75	6	3189	4229	gnl|PID|e209890	NAD alcohol dehydrogenase [Bacillus subtilis]	55	44	1041
108	10	10488	9820	gnl|PID|e324997	hypothetical protein [Bacillus subtilis]	55	36	669
113	12	12273	13037	gnl|PID|e311496	unknown [Bacillus subtilis]	55	34	765
113	13	13007	13945	gi|1573423	1-phosphofructokinase (fruK) [Haemophilus influenzae]	55	39	939
126	5	6764	5907	gi|1790131	(AE000446) hypothetical 29.7 kD protein in ibpA-gyrB intergenic region [Escherichia coli]	55	37	858

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
129	3	2719	902	|gnl|PID|d101425	Pz-peptidase (Bacillus licheniformis)	55	35	1818
138	3	2593	1610	|gi|142833	ORF2 (Bacillus subtilis)	55	37	984
140	6	6916	5633	|gnl|PID|d100964	homologue of hypothetical protein in a rapamycin synthesis gene cluster of Streptomyces hygroscopicus (Bacillus subtilis)	55	26	1284
147	3	3854	2136	|gi|472330	dihydrolipoamide dehydrogenase (Clostridium magnum)	55	39	1719
147	10	10204	8921	|gnl|PID|e73078	dihydroorotase (Lactobacillus leichmannii)	55	38	1284
148	5	3430	4119	|gi|290572	peripheral membrane protein U (Escherichia coli)	55	29	690
148	6	4171	4650	|gi|695769	transposase (Xanthobacter autotrophicus)	55	37	480
149	14	12564	11650	|gnl|PID|d101329	YqjG (Bacillus subtilis)	55	32	915
156	3	1113	550	|gi|2314496	(AE000634) conserved hypothetical integral membrane protein (Helicobacter pylori)	55	34	564
159	10	6625	5897	|gi|290533	similar to E. coli ORF adjacent to suc operon; similar to gntR class of regulatory proteins (Escherichia coli)	55	29	729
164	3	1784	2332	|gnl|PID|e255118	hypothetical protein (Bacillus subtilis)	55	37	549
164	5	2772	3521	|gi|40348	put. resolvase Tnp I (AA 1 - 284) (Bacillus thuringiensis)	55	35	750
164	11	7428	7216	|gnl|PID|e249407	unknown (Mycobacterium tuberculosis)	55	38	213
167	5	3860	3345	|gi|535052	involved in protein secretion (Bacillus subtilis)	55	28	516
186	5	2880	2563	|gi|606080	ORF_o290; Geneplot suggests frameshift linking to o267, not found (Escherichia coli)	55	35	318
189	8	4311	5396	|gnl|PID|e183450	hypothetical EcsB protein (Bacillus subtilis)	55	32	1086
192	5	3270	3079	|gi|1196504	vitellogenin convertase (Aedes aegypti)	55	38	192
195	2	2454	1384	|gi|1574693	transferase, peptidoglycan synthesis (murG) (Haemophilus influenzae)	55	33	1071
198	4	3013	2471	|gnl|PID|e313074	hypothetical protein (Bacillus subtilis)	55	29	543
214	1	373	744	|gnl|PID|d101741	transposase (Synechocystis sp.)	55	33	372
219	2	1115	456	|gi|288301	ORF2 gene product (Bacillus megaterium)	55	30	660
263	7	3742	3443	|gi|18137	cgcr-4 product (Chlamydomonas reinhardtii)	55	48	300
285	1	2	829	|gnl|PID|d100974	unknown (Bacillus subtilis)	55	40	828
286	1	650	249	|gi|396844	ORF (18 kDa) (Vibrio cholerae)	55	31	402
297	2	1229	1696	|gi|150848	prtC (Porphyromonas gingivalis)	55	39	468

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
309	2	218	982	gi|1574491	hypothetical [Haemophilus influenzae]	55	35	765
328	2	646	224	gi|571500	prohibitin [Saccharomyces cerevisiae]	55	27	423
330	1	1340	474	gi|396397	soxS [Escherichia coli]	55	29	867
364	3	2538	1546	gi|393394	Tb-291 membrane associated protein [Trypanosoma brucei subgroup]	55	36	993
368	3	941	105	gi|160671	S antigen precursor [Plasmodium falciparum]	55	40	837
3	5	4604	3624	gi|2293176	(AF008220) signal transduction protein kinase [Bacillus subtilis]	54	26	981
9	11	7746	7246	gi|1146245	putative [Bacillus subtilis]	54	38	501
38	24	16213	17937	gi|1480429	putative transcriptional regulator [Bacillus stearothermophilus]	54	27	1725
40	8	5076	4882	gi|39989	methionyl-tRNA synthetase [Bacillus stearothermophilus]	54	35	195
43	4	3980	2367	gnl|PID|e148611	ABC transporter [Lactobacillus helveticus]	54	25	1614
52	10	10844	12103	gi|1762962	FemA [Staphylococcus simulans]	54	29	1260
57	1	3	512	gi|558177	endo-1,4-beta-xylanase [Cellulomonas fimi]	54	36	510
58	3	4749	4246	gnl|PID|d101237	hypothetical [Bacillus subtilis]	54	29	504
71	7	10684	11703	gi|510255	orf3 [Escherichia coli]	54	31	1020
71	20	27546	27737	gi|202543	serotonin receptor [Rattus norvegicus]	54	31	192
72	2	844	1098	gi|148613	srnB gene product [Plasmid F]	54	37	255
72	7	7438	6695	gi|1196496	recombinase [Moraxella bovis]	54	38	744
74	10	14043	13465	gi|1200342	ORF 3 gene product [Bradyrhizobium japonicum]	54	32	579
74	12	16483	15995	gi|2317798	maturase-related protein [Pseudomonas alcaligenes]	54	30	489
86	3	2877	2155	gi|46988	orf9.6 possibly encodes the O unit polymerase [Salmonella enterica]	54	34	723
89	5	4433	3921	gi|147211	phnO protein [Escherichia coli]	54	41	513
90	1	3	464	gi|2317798	maturase-related protein [Pseudomonas alcaligenes]	54	30	462
96	10	8058	8510	gnl|PID|d102015	(AB001488) SIMILAR TO SALMONELLA TYPHIMURIUM SLYY GENE REQUIRED FOR SURVIVAL IN MACROPHAGE. [Bacillus subtilis]	54	32	453
97	6	4662	3604	gi|1591394	transketolase'' [Methanococcus jannaschii]	54	30	1059
106	11	10406	12010	gi|606286	ORF_o637 [Escherichia coli]	54	32	1605
147	8	8663	7404	gnl|PID|d101615	ORF_ID:o31987; similar to [SwissProt Accession Number P37340] [Escherichia coli]	54	35	1260

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
171	4	2477	3223	|gi|1439528	|EIIC-man (Lactobacillus curvatus)	54	36	747
174	2	2068	1787	|gnl|PID|d100518	|motor protein (Homo sapiens)	54	35	282
188	1	526	1188	|gnl|PID|e250352	|unknown (Mycobacterium tuberculosis)	54	31	663
198	5	3582	2884	|gnl|PID|e313074	|hypothetical protein (Bacillus subtilis)	54	33	699
207	1	1	1641	|gnl|PID|d101813	|hypothetical protein (Synechocystis sp.)	54	24	1641
210	1	2	655	|gi|2293206	|(AF008220) YtmP (Bacillus subtilis)	54	29	654
225	2	966	2357	|gnl|PID|e330194	|R11H6.1 (Caenorhabditis elegans)	54	39	1392
241	1	1681	347	|gnl|PID|d101813	|hypothetical protein (Synechocystis sp.)	54	26	1335
263	2	907	1395	|gnl|PID|d101886	|transposase (Synechocystis sp.)	54	30	489
263	6	3450	2977	|gi|160671	|S antigen precursor (Plasmodium falciparum)	54	47	474
277	3	2517	1363	|gi|1196926	|unknown protein (Streptococcus mutans)	54	30	1155
307	1	828	4	|gi|2293198	|(AF008220) YtgP (Bacillus subtilis)	54	28	825
325	1	19	768	|gi|2182507	|(AE000083) Y41H (Rhizobium sp. NGR234)	54	37	750
332	2	898	590	|gi|1591815	|ADP-ribosylglycohydrolase (draG) (Methanococcus jannaschii)	54	32	309
385	4	240	479	|gi|530878	|amino acid feature: N-glycosylation sites, aa 41 .. 43, 46 .. 48, 51 .. 53, 72 .. 74, 107 .. 109, 128 .. 130, 132 .. 134, 158 .. 160, 163 .. 165; amino acid feature: Rod protein domain, aa 169 .. 340; amino acid feature: globular protein domai	54	49	240
7	25	19702	19493	|gnl|PID|e255111	|hypothetical protein (Bacillus subtilis)	53	32	210
23	3	2497	2033	|gnl|PID|d102015	|(AB001488) SIMILAR TO SALMONELLA TYPHIMURIUM SLYY GENE REQUIRED FOR SURVIVAL IN MACROPHAGE. (Bacillus subtilis)	53	25	465
29	11	9042	10121	|gi|143331	|alkaline phosphatase regulatory protein (Bacillus subtilis)	53	31	1080
33	3	1479	1009	|pir|S10655|S106	|hypothetical protein X - Pyrococcus woesei (fragment)	53	33	471
36	6	4583	5134	|gnl|PID|e316029	|unknown (Mycobacterium tuberculosis)	53	30	552
38	14	8521	8898	|gi|580904	|homologous to E coli rnpA (Bacillus subtilis)	53	30	378
52	7	7007	8686	|gi|1377831	|unknown (Bacillus subtilis)	53	29	1680
54	17	17555	19564	|gi|666069	|orf2 gene product (Lactobacillus leichmannii)	53	36	2010
56	1	1	681	|gi|1592266	|restriction modification system S subunit (Methanococcus jannaschii)	53	32	681

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
57	10	9431	8487	|gi|1788543	(AE000310) f351; Residues 1-121 are 100 pct identical to YOJL_ECOLI SW: P33944 (122 aa) and aa 152-351 are 100 pct identical to YOJK_ECOLI SW: P33943 (Escherichia coli)	53	31	945
61	1	429	4	|gnl|PID|e236467	B0024.12 (Caenorhabditis elegans)	53	33	426
71	1	5772	4	|gi|393394	(Tb-291 membrane associated protein (Trypanosoma brucei subgroup)	53	33	5769
72	3	894	2840	|gi|2293178	(AF008220) YtsD (Bacillus subtilis)	53	27	1947
73	14	9793	9212	|gi|1778556	(putative cobalamin synthesis protein (Escherichia coli)	53	32	582
88	7	5217	4342	|gi|2098719	(putative fimbrial-associated protein (Actinomyces naeslundii)	53	38	876
93	5	2395	1688	|gi|563366	(gluconate oxidoreductase (Gluconobacter oxydans)	53	33	708
96	9	6632	7762	|gi|517204	(ORF1, putative 42 kDa protein (Streptococcus pyogenes)	53	42	1131
108	8	7629	8600	|gi|149581	(maturation protein (Lactobacillus paracasei)	53	32	972
128	9	6412	6972	|gnl|PID|e317237	(unknown (Mycobacterium tuberculosis)	53	36	561
128	12	8429	9253	|gi|311070	(pentraxin fusion protein (Xenopus laevis)	53	31	825
148	1	3	950	|pir|A61607|A616	(probable hemolysin precursor - Streptococcus agalactiae (strain 74-360)	53	36	948
163	2	2162	3022	|gi|1755150	(nocturnin (Xenopus laevis)	53	30	861
171	3	2304	2624	|gi|1732200	(PTS permease for mannose subunit IIPMan (Vibrio furnissii)	53	32	321
182	5	3785	3051	|gnl|PID|d100572	(unknown (Bacillus subtilis)	53	35	735
209	3	2948	1935	|gi|1778505	(ferric enterobactin transport protein (Escherichia coli)	53	28	1014
218	5	3884	2406	|gi|40162	(murE gene product (Bacillus subtilis)	53	34	1479
250	3	473	790	|gnl|PID|e334776	(YIbH protein (Bacillus subtilis)	53	30	318
275	3	1	1611	|gnl|PID|d101314	(YqeW (Bacillus subtilis)	53	35	1611
332	1	544	2	|gi|409286	(bmrU (Bacillus subtilis)	53	31	543
2	2	2543	3445	|gnl|PID|e233879	(hypothetical protein (Bacillus subtilis)	52	39	903
3	22	22402	23376	|gi|38969	(lacF gene product (Agrobacterium radiobacter)	52	36	975
5	3	8094	2356	|gnl|PID|e324915	(IgA1 protease (Streptococcus sanguis)	52	32	5739
22	26	19961	20212	|gi|152901	(ORF 3 (Spirochaeta aurantia)	52	35	252
22	31	23140	24666	|gi|289262	(comE ORF3 (Bacillus subtilis)	52	32	1527
27	6	5397	4801	|gi|39573	(P20 (AA 1-178) (Bacillus licheniformis)	52	35	597

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
35	10	8604	7357	gi|508241	putative O-antigen transporter [Escherichia coli]	52	27	1248
45	4	4801	3662	gnl|PID|d102243	(AB005554) homologs are found in E. coli and H. influenzae; see SWISS_PROT ACC#: P42100 [Bacillus subtilis]	52	36	1140
48	18	14385	13726	gnl|PID|e205174	orf2 [Lactobacillus helveticus]	52	25	660
49	4	5321	5755	gi|2317740	(AF013987) nitrogen regulatory IIA protein [Vibrio cholerae]	52	19	435
54	4	2773	4668	gi|1500472	M. jannaschii predicted coding region MJ1577 [Methanococcus jannaschii]	52	36	1896
54	6	5250	4969	gi|2182453	(AE000079) Y4iO [Rhizobium sp. NGR234]	52	40	282
66	6	8400	6955	gi|43140	TrkG protein [Escherichia coli]	52	30	1446
71	26	30659	31312	gnl|PID|e314993	unknown [Mycobacterium tuberculosis]	52	23	654
75	2	1673	1035	gnl|PID|d102271	(AB001683) FarA [Streptomyces sp.]	52	27	639
81	3	1439	2893	gnl|PID|e311458	rhamnulose kinase [Bacillus subtilis]	52	32	1455
81	8	4987	5781	gi|147403	mannose permease subunit II-P-Man [Escherichia coli]	52	37	795
83	21	20687	21853	gi|143365	phosphoribosyl aminoimidazole carboxylase II (PUR-K; ttg start codon) [Bacillus subtilis]	52	37	1167
86	6	5785	4592	gi|1276879	EpsF [Streptococcus thermophilus]	52	26	1194
86	20	19390	17861	gi|454844	ORF 3 [Schistosoma mansoni]	52	26	1530
96	13	10540	9659	gi|288299	ORF1 gene product [Bacillus megaterium]	52	33	882
111	1	2	2026	gi|148309	cytolysin B transport protein [Enterococcus faecalis]	52	27	2025
112	2	1457	2167	gi|471234	orf1 [Haemophilus influenzae]	52	33	711
118	3	2931	2365	bbs|151233	Mip=24 kda macrophage infectivity potentiator protein [Legionella pneumophila, Philadelphia-1, Peptide, 184 aa] [Legionella pneumophila]	52	33	567
122	9	5646	5951	gi|8214	myosin heavy chain [Drosophila melanogaster]	52	36	306
122	11	6159	6374	gi|434025	dihydrolipoamide acetyltransferase [Pelobacter carbinolicus]	52	52	216
134	6	4880	6313	gi|151733	M protein trans-acting positive regulator [Streptococcus pyogenes]	52	43	1434
135	3	1238	2716	gnl|PID|e245024	unknown [Mycobacterium tuberculosis]	52	35	1479
141	3	1681	2319	gnl|PID|d100573	unknown [Bacillus subtilis]	52	32	639
161	4	2562	5024	gi|1146243	22 4% identity with Escherichia coli DNA-damage inducible protein . . . putative [Bacillus subtilis]	52	36	2463
173	2	968	183	gi|1215693	putative orf; GT9_orf434 [Mycoplasma pneumoniae]	52	30	786

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
198	6	4400	3567	|gnl|PID|e313010	|hypothetical protein (Bacillus subtilis)	52	26	834
210	12	8844	9107	|gi|497647	|DNA gyrase subunit B (Mycoplasma genitalium)	52	38	264
214	10	5264	5431	|gi|550697	|envelope protein (Human immunodeficiency virus type 1)	52	36	168
225	1	15	884	|gi|1552773	|hypothetical (Escherichia coli)	52	34	870
230	1	39	362	|gnl|PID|d100582	|unknown (Bacillus subtilis)	52	28	324
287	1	871	2	|gnl|PID|e335028	|protease/peptidase (Mycobacterium leprae)	52	29	870
363	2	1305	4	|gi|393394	|Tb-291 membrane associated protein (Trypanosoma brucei subgroup)	52	32	1302
23	2	2048	1173	|gnl|PID|e254943	|unknown (Mycobacterium tuberculosis)	51	30	876
29	3	742	1521	|gi|929900	|5'-methylthioadenosine phosphorylase (Sulfolobus solfataricus)	51	31	780
45	1	410	1597	|gi|1877429	|integrase (Streptococcus pyogenes phage TI2)	51	32	1188
48	26	19227	18946	|gi|2314455	|(AE000633) transcriptional regulator (tenA) (Helicobacter pylori)	51	33	282
73	5	4276	4016	|gi|474177	|alpha-D-1,4-glucosidase (Staphylococcus xylosus)	51	31	261
81	11	8935	12057	|gi|311070	|pentraxin fusion protein (Xenopus laevis)	51	31	3123
83	5	1195	1986	|gnl|PID|d101316	|YqfI (Bacillus subtilis)	51	33	792
98	10	7531	8538	|gi|41500	|ORF 3 (AA 1-352); 38 kD (put. ftsX) (Escherichia coli)	51	28	1008
111	6	3908	5173	|gi|466882	|pps1; B1496_C2_189 (Mycobacterium leprae)	51	27	1266
124	1	326	57	|gi|2191168	|(AF007270) contains similarity to myosin heavy chain (Arabidopsis thaliana)	51	32	270
129	10	7286	6816	|gi|1046241	|orf14 (Bacteriophage HP1)	51	30	471
143	3	4963	3983	|gi|1354935	|probable copper-transporting atpase (Escherichia coli)	51	26	981
148	15	11359	10226	|gi|2293256	|(AF008220) putative hippurate hydrolase (Bacillus subtilis)	51	36	1134
149	8	6003	7313	|gi|1633572	|Herpesvirus saimiri ORF73 homolog (Kaposi's sarcoma-associated herpes-like virus)	51	21	1311
151	9	12092	11550	|gnl|PID|e281580	|hypothetical 40.7 kd protein (Bacillus subtilis)	51	34	543
159	6	2555	3208	|gi|146944	|CMP-N-acetylneuraminic acid synthetase (Escherichia coli)	51	36	654
174	1	1797	4	|gi|1773166	|probable copper-transporting atpase (Escherichia coli)	51	28	1794
265	4	2231	1773	|gnl|PID|e256400	|anti-P.falciparum antigenic polypeptide (Saimiri sciureus)	51	18	459
277	2	643	1311	|pir|S32915|S329	|pilD protein - Neisseria gonorrhoeae	51	33	669

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
350	1	890	3	gi|290509	o307 (Escherichia coli)	51	30	888
363	4	1228	4485	gi|1707247	partial CDS (Caenorhabditis elegans)	51	23	3258
367	1	1701	4	gi|393394	Tb-291 membrane associated protein (Trypanosoma brucei subgroup)	51	32	1698
15	5	5174	4497	gnl|PID|e58151	F3 (Bacillus subtilis)	50	38	678
16	4	2220	2582	gnl|PID|e325010	hypothetical protein (Bacillus subtilis)	50	29	363
19	5	2591	4159	gi|1552733	similar to voltage-gated chloride channel protein (Escherichia coli)	50	30	1569
25	4	2701	1997	gi|887849	ORF_f219 (Escherichia coli)	50	27	705
35	1	211	417	gnl|PID|e236697	unknown (Saccharomyces cerevisiae)	50	33	207
39	4	3416	5152	gnl|PID|d100974	unknown (Bacillus subtilis)	50	27	1737
51	7	4000	5181	gi|1592027	carbamoyl-phosphate synthase, pyrimidine-specific, large subunit (Methanococcus jannaschii)	50	27	1182
51	9	7179	8303	gi|1591847	type I restriction-modification enzyme, S subunit (Methanococcus jannaschii)	50	28	1125
52	8	8740	9534	gi|144297	acetyl esterase (XynC) (Caldocellum saccharolyticum)	50	34	795
52	16	16591	15770	gi|2108229	basic surface protein (Lactobacillus fermentum)	50	34	822
57	7	6031	6336	gi|2275264	60S ribosomal protein L7B (Schizosaccharomyces pombe)	50	40	306
71	23	29348	28383	gnl|PID|d101328	YqjA (Bacillus subtilis)	50	30	966
86	12	11155	10769	gnl|PID|e324964	hypothetical protein (Bacillus subtilis)	50	24	387
93	2	1205	330	gi|1066016	similar to Escherichia coli pyruvate, water dikinase, Swiss-Prot Accession Number P23538 (Pyrococcus furiosus)	50	24	876
96	5	1673	2959	gnl|PID|e322433	gamma-glutamylcysteine synthetase (Brassica juncea)	50	29	1287
98	2	218	1171	gi|151110	leucine-, isoleucine-, and valine-binding protein (Pseudomonas aeruginosa)	50	30	954
103	4	3303	2785	gi|154330	O-antigen ligase (Salmonella typhimurium)	50	31	519
115	5	6480	5980	gi|895747	putative cel operon regulator (Bacillus subtilis)	50	26	501
129	11	7559	7305	gi|1216475	skeletal muscle ryanodine receptor (Homo sapiens)	50	32	255
129	13	8192	7965	gi|152271	319-kDA protein (Rhizobium meliloti)	50	30	228
151	5	7634	6819	gi|40348	put. resolvase Tnp I (AA 1 - 284) (Bacillus thuringiensis)	50	35	816
153	1	1	597	gnl|PID|d102015	(AB0001488) SIMILAR TO NITROREDUCTASE. (Bacillus subtilis)	50	29	597

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
155	5	5986	5432	gi|1276880	EpsG (Streptococcus thermophilus)	50	28	555
160	9	7390	6323	gi|1786983	(AE000179) o331; 92 pct identical to the 333 aa hypothetical protein YBHE_ECOLI SW: P52697; 26 pct identical (7 gaps) to 167 residues of the 373 aa protein MLE_TRICU SW: P46057; SW: P52697 (Escherichia coli)	50	30	1068
163	6	7396	8091	gnl|PID|d101313	YqeN (Bacillus subtilis)	50	22	696
167	6	5232	3940	gi|413926	ipa-2r gene product (Bacillus subtilis)	50	27	1293
169	2	807	130	gnl|PID|e304540	endolysin (Bacteriophage Bastille)	50	35	678
171	5	3168	4025	gi|606080	ORF_o290; Geneplot suggests frameshift linking to o267, not found (Escherichia coli)	50	27	858
210	11	8151	8414	gi|330038	HRV 2 polyprotein (Human rhinovirus)	50	25	264
164	1	1538	135	gi|393396	Tb-292 membrane associated protein (Trypanosoma brucei subgroup)	50	31	1404
10	7	5911	5090	gi|144859	ORF B (Clostridium perfringens)	49	24	822
26	5	10754	9768	gi|142440	ATP-dependent nuclease (Bacillus subtilis)	49	31	987
66	7	9777	8398	gi|414170	trkA gene product (Methanosarcina mazeii)	49	26	1380
77	6	5364	4648	gnl|PID|e285322	RecX protein (Mycobacterium smegmatis)	49	28	717
82	13	12689	13249	gnl|PID|e255091	hypothetical protein (Bacillus subtilis)	49	20	561
93	9	4866	4531	gi|40067	X gene product (Bacillus sphaericus)	49	26	336
112	5	4019	4948	gi|1574380	lic-1 operon protein (licB) (Haemophilus influenzae)	49	27	930
129	7	6058	4949	gnl|PID|e267587	Unknown (Bacillus subtilis)	49	35	1110
135	5	3875	4438	gi|39573	P20 (AA 1-178) (Bacillus licheniformis)	49	25	564
154	2	1423	1953	gnl|PID|d101102	regulatory components of sensory transduction system (Synechocystis sp.)	49	29	531
156	5	2878	1637	gnl|PID|d101732	hypothetical protein (Synechocystis sp.)	49	25	1242
173	5	3500	2940	gi|490324	LORF X gene product (unidentified)	49	30	561
182	1	1057	2	gi|331002	first methionine codon in the ECLF1 ORF (Saimiriine herpesvirus 2)	49	25	1056
192	6	5352	3667	gi|2394472	(AF024499) contains similarity to homeobox domains (Caenorhabditis elegans)	49	23	1686
253	4	1129	1350	gi|531116	SIR4 protein (Saccharomyces cerevisiae)	49	23	222
277	1	600	136	gi|396844	ORF (18 kDa) (Vibrio cholerae)	49	32	465
327	3	1435	887	gi|733524	phosphatidylinositol-4,5-diphosphate 3-kinase (Dictyostelium discoideum)	49	24	549

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
365	3	1436	132	|gi|393394	|Tb-291 membrane associated protein (Trypanosoma brucei subgroup)	49	31	1305
33	7	4461	3277	|gi|145644	|codes for a protein of unknown function (Escherichia coli)	48	26	1185
40	2	652	1776	|gnl|PID|e290649	|ornithine decarboxylase (Nicotiana tabacum)	48	29	1125
67	4	1377	2384	|gi|1772652	|2-keto-3-deoxygluconate kinase (Haloferax alicantei)	48	30	1008
74	2	4269	3871	|gi|2182678	|(AE000101) Y4vJ (Rhizobium sp. NGR234)	48	27	399
81	2	1326	541	|gi|153672	|lactose repressor (Streptococcus mutans)	48	33	786
81	4	2981	3646	|gi|146042	|fuculose-1-phosphate aldolase (fucA) (Escherichia coli)	48	30	666
97	1	602	51	|gi|153794	|rgg (Streptococcus gordonii)	48	29	552
110	1	1	3132	|gi|1381114	|prtB gene product (Lactobacillus delbrueckii)	48	23	3132
131	5	2914	2147	|gnl|PID|e183811	|Acyl-ACP thioesterase (Brassica napus)	48	27	768
133	4	3494	2628	|gnl|PID|e261988	|putative ORF (Bacillus subtilis)	48	27	867
139	6	4231	4599	|gi|1049388	|ZK470.1 gene product (Caenorhabditis elegans)	48	23	369
139	8	5036	5665	|gi|1022725	|unknown (Staphylococcus haemolyticus)	48	29	630
140	12	11936	11007	|gnl|PID|d102049	|H. influenzae, ribosomal protein alanine acetyltransferase; P44305 (189) (Bacillus subtilis)	48	27	930
146	9	5670	4654	|gi|1591731	|melvalonate kinase (Methanococcus jannaschii)	48	24	1017
161	3	1280	2374	|gnl|PID|d101578	|Collagenase precursor (EC 3.4.-.-). (Escherichia coli)	48	24	1095
172	11	10581	11048	|gnl|PID|d101132	|hypothetical protein (Synechocystis sp.)	48	27	468
182	4	2930	2586	|gi|40067	|X gene product (Bacillus sphaericus)	48	37	345
210	15	10786	11196	|sp|P13940|LE29_	|LATE EMBRYOGENESIS ABUNDANT PROTEIN D-29 (LEA D-29).	48	30	411
214	12	6231	6482	|gi|40389	|non-toxic components (Clostridium botulinum)	48	26	252
221	1	704	3	|gi|1573364	|H influenzae predicted coding region HI0392 (Haemophilus influenzae)	48	27	702
227	2	647	3928	|gi|1673693	|(AE000005) Mycoplasma pneumoniae, C09_orf718 Protein (Mycoplasma pneumoniae)	48	30	3282
251	2	480	758	|gnl|PID|e236697	|unknown (Saccharomyces cerevisiae)	48	31	279
363	3	1874	1122	|gi|18137	|cgcr-4 product (Chlamydomonas reinhardtii)	48	40	753
389	1	505	2	|gi|18137	|cgcr-4 product (Chlamydomonas reinhardtii)	48	38	504
3	21	20879	22258	|gnl|PID|e264778	|putative maltose-binding pootein (Streptomyces coelicolor)	47	33	1380

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
6	4	4089	4658	gi|39573	P20 (AA 1-178) [Bacillus licheniformis]	47	23	570
15	3	3736	1760	gnl|PID|d100572	unknown [Bacillus subtilis]	47	25	1977
35	15	14516	13263	gi|1773351	Cap5L [Staphylococcus aureus]	47	20	1254
51	6	3547	4002	pir|A37024|A370	32K antigen precursor - Mycobacterium tuberculosis	47	38	456
55	8	10154	9273	gi|39848	UJ [Bacillus subtilis]	47	26	882
92	4	1753	3276	gnl|PID|e280611	PCPC [Streptococcus pneumoniae]	47	35	1524
127	9	5589	5386	gi|1786458	(AE000134) f120; This 120 aa orf is 76 pct identical (0 gaps) to 42 residues of an approx. 48 aa protein Y127_HAEIN SW: P43949 [Escherichia coli]	47	32	204
130	2	1232	1759	gnl|PID|e266555	unknown [Mycobacterium tuberculosis]	47	23	528
140	4	4951	3542	gnl|PID|d100964	homologue of hypothetical protein in a rapamycin synthesis gene cluster of Streptomyces hygroscopicus [Bacillus subtilis]	47	24	1410
151	4	6814	6200	gi|1522674	M. jannaschii predicted coding region MJECL41 [Methanococcus jannaschii]	47	27	615
157	3	803	1174	gnl|PID|d101320	Yqg2 [Bacillus subtilis]	47	25	372
178	5	3267	2155	gi|2367190	(AE000390) o334; sequence change joins ORFs ygjR & ygjS from earlier version (YGJR_ECOLI SW: P42599 and YGJS_ECOLI SW: P42600) [Escherichia coli]	47	30	1113
273	1	2	1549	gnl|PID|e254973	autolysin sensor kinase [Bacillus subtilis]	47	32	1548
300	2	880	644	gi|1835755	zinc finger protein Png-1 [Mus musculus]	47	22	237
54	14	14182	12638	pir|S43609|S436	rofA protein - Streptococcus pyogenes	46	24	1545
88	1	2	1018	gnl|PID|e223891	xylose repressor [Anaerocellum thermophilum]	46	27	1017
96	7	4553	5860	gnl|PID|d101652	ORF_ID:o34785; similar to [SwissProt Accession Number P45272] [Escherichia coli]	46	23	1308
112	1	1127	3	gi|2209215	(AF004325) putative oligosaccharide repeat unit transporter [Streptococcus pneumoniae]	46	24	1125
122	13	7308	7982	gi|1054776	hr44 gene product [Homo sapiens]	46	34	675
127	14	9198	8125	gi|1469286	afuA gene product [Actinobacillus pleuropneumoniae]	46	28	1074
132	4	7093	6197	gi|153794	rgg [Streptococcus gordonii]	46	26	897
140	8	8220	7723	gi|1235795	pullulanase [Thermoanaerobacterium thermosulfurigenes]	46	21	498
140	9	9205	8315	gi|407878	leucine rich protein [Streptococcus equisimilis]	46	27	891

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
162	1	1	1125	gi|1143209	ORF7; Method: conceptual translation supplied by author [Shigella sonnei]	46	25	1125
199	1	1	585	gi|1947171	(AF000299) No definition line found [Caenorhabditis elegans]	46	28	585
223	3	1971	1477	sp|P02562|MYSS_	MYOSIN HEAVY CHAIN, SKELETAL MUSCLE (FRAGMENTS).	46	27	495
232	2	760	1608	gi|1016112	ycf38 gene product [Cyanophora paradoxa]	46	28	849
292	1	687	220	gi|1673744	(AE000011) Mycoplasma pneumoniae, cytidine deaminase; similar to GenBank Accession Number C53312, from H. pirum [Mycoplasma pneumoniae]	46	29	468
30	8	5843	6472	gi|1788049	(AE000270) o235; This 235 aa orf is 29 pct identical (10 gaps) to 198 residues of an approx. 216 aa protein YTXB_BACSU SW: P06568 [Escherichia coli]	45	24	630
48	6	3461	3868	gi|722339	unknown [Acetobacter xylinum]	45	29	408
60	1	307	2	gi|1699079	coded for by C. elegans cDNA yk41h4.3; coded for by C. elegans cDNA yk148g10.5; coded for by C. elegans cDNA yk152g5.5; coded for by C. elegans cDNA yk59a10.5; coded for by C. elegans cDNA yk41h4 5, coded for by C. elegans cDNA cm20g10; coded	45	36	306
72	16	14371	14874	gi|1321900	NADH dehydrogenase (ubiquinone) [Artemia franciscana]	45	25	504
99	7	9158	7941	gi|152192	mutation causes a succinoglucan-minus phenotype; ExoQ is atransmembrane protein; third gene of the exoYFQ operon;; putative [Rhizobium meliloti]	45	28	1218
127	12	7046	6606	bbs|153689	HitB-iron utilization protein [Haemophilus influenzae, type b, DL42, NTHI TN106, Peptide, 506 aa] [Haemophilus influenzae]	45	24	441
137	5	1561	2619	gi|472921	v-type Na-ATPase [Enterococcus hirae]	45	33	1059
209	1	774	364	gi|304141	restriction endonuclease beta subunit [Bacillus coagulans]	45	28	411
314	1	604	2	gi|1480457	latex allergen [Hevea brasiliensis]	45	31	603
20	18	19782	20288	gi|433942	ORF [Lactococcus lactis]	44	26	507
87	8	7030	6452	gi|537207	ORF_f277 [Escherichia coli]	44	26	579
166	5	4909	4037	gnl|PID|e308082	membrane transport protein [Bacillus subtilis]	44	25	873
247	1	818	75	gnl|PID|d100718	ORF1 [Bacillus sp.]	44	20	744
32	3	1885	3876	gi|2351768	PspA [Streptococcus pneumoniae]	43	24	1992
36	17	15467	18256	gi|1045739	M. genitalium predicted coding region MG064 [Mycoplasma genitalium]	43	26	2790
54	15	14656	17343	gi|520541	penicillin-binding proteins 1A and 1B [Bacillus subtilis]	43	27	2688
67	2	696	1352	gi|536934	yjcA gene product [Escherichia coli]	43	29	657
139	2	2416	338	gi|396400	similar to eukaryotic Na+/H+ exchangers [Escherichia coli]	43	24	2079

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
298	1	3	809	gi|413972	ipa-48r gene product (Bacillus subtilis)	43	24	807
387	1	47	427	gi|2315652	(AF016669) No definition line found (Caenorhabditis elegans)	43	30	381
185	4	4221	3127	gi|2182399	(AE000073) Y4fP (Rhizobium sp. NGR234)	41	25	1095
340	1	582	70	gnl|PID|e218681	CDP-diacylglycerol synthetase (Arabidopsis thaliana)	41	20	513
363	6	4205	1914	gi|1256742	R27-2 protein (Trypanosoma cruzi)	41	27	2292
368	2	2	943	gi|21783	LMW glutenin (AA 1-356) (Triticum aestivum)	41	34	942
155	3	4489	2861	gi|42023	member of ATP-dependent transport family, very similar to mdr proteins and hemolysin B, export protein (Escherichia coli)	40	18	1629
365	2	95	1438	gi|1633572	Herpesvirus saimiri ORF73 homolog (Kaposi's sarcoma-associated herpes-like virus)	40	21	1344
1	3	2979	3860	gnl|PID|d101908	hypothetical protein (Synechocystis sp.)	39	26	882
1	5	3814	4647	gnl|PID|d101961	hypothetical protein (Synechocystis sp.)	39	19	834
26	6	14035	10724	gi|142439	ATP-dependent nuclease (Bacillus subtilis)	38	20	3312
47	1	3	4916	gi|632549	NF-180 (Petromyzon marinus)	36	23	4914

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
1	4	3428	3009
1	6	4611	4964
3	2	818	994
3	3	1182	1574
3	7	5382	6497
3	25	25046	25396
3	26	25625	26317
6	2	1519	1689
6	14	12875	12618
6	15	13215	12841
6	18	15977	15390
7	12	9955	9419
7	13	10161	9910
8	6	3915	4280
9	9	6024	5704
10	8	6909	6298
10	9	7136	6888
10	11	7968	7672
12	1	1140	4
12	3	1779	1456
14	2	1913	1434
16	1	1	243
16	5	5675	3087
17	1	324	34
17	3	1451	1050
17	9	4890	4465
20	14	14544	15893

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
21	3	3359	2589
21	5	4802	4482
22	21	17099	17362
22	25	19467	19982
22	33	25540	25764
22	35	26388	26218
22	36	26382	27572
23	7	6655	6032
23	8	7132	6653
24	1	36	518
25	5	3009	2641
27	4	4819	4223
27	5	4789	4956
28	5	3017	1797
28	8	4272	3850
28	10	5028	4597
28	11	5746	5072
29	7	5596	4919
29	8	5039	5518
29	9	5595	8207
30	9	6511	6263
31	6	2664	2344
32	5	5203	5538
33	8	5327	4668
34	10	8024	7740
34	12	9360	8641
34	13	9667	9377

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
34	18	13104	11902
35	11	9688	8588
35	12	11073	9670
36	2	334	1041
36	12	11120	10893
36	13	10993	11388
36	15	12172	14595
38	7	4269	4577
38	8	4480	5001
38	10	5517	5711
38	17	10732	11376
40	3	1728	3143
43	1	172	5
43	7	8884	8732
43	8	9568	9071
44	4	4831	6831
45	3	3204	3665
46	4	3875	3468
46	7	6074	7081
48	5	3196	3582
48	8	4579	4229
48	11	9323	8922
48	16	13042	12494
48	20	16342	15764
48	24	17971	18351
48	30	21979	21776
49	1	209	3

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
50	4	3307	2672
51	5	3239	3598
52	11	12146	12883
54	7	5588	5187
54	8	6013	5459
54	9	6004	6210
54	16	17685	17506
55	9	10515	10123
55	12	11947	12141
56	3	935	1387
56	4	1496	1939
57	3	1624	2130
57	4	2100	2501
58	6	7541	7335
59	1	2	430
59	4	2416	2736
59	5	2734	3063
59	8	4743	5549
59	9	5459	5929
60	6	5741	6451
61	3	2395	1772
61	5	3316	3176
64	1	2722	2
66	2	1180	3147
66	8	9082	9495
67	3	1343	1182
69	2	1165	980

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
70	5	4059	3922
70	6	4215	4057
70	9	5268	5504
71	15	20351	21901
71	16	21859	22338
71	19	26204	27556
72	9	8458	8081
73	4	3815	4216
73	6	4214	4582
73	7	4369	4773
73	10	7183	6428
73	15	9462	9668
76	1	524	195
76	2	867	535
76	11	8602	9210
80	6	7924	8109
81	1	244	2
81	10	6631	8931
83	4	1872	1150
83	17	16810	16460
84	3	4464	2929
86	2	2147	1092
86	4	3606	2875
86	19	16767	17114
87	5	5326	5000
87	7	6459	6001
87	9	7224	7006

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
87	18	17930	17670
87	19	18275	17928
88	2	1619	1840
88	4	2711	2878
88	9	6252	6016
89	3	2634	1621
89	9	7171	6868
90	2	899	2395
90	3	1143	952
91	3	2959	3141
91	4	3170	3691
91	6	4253	4573
93	1	391	2
93	6	2648	2379
93	8	4533	3712
96	1	3	182
96	2	904	632
96	3	1407	1147
96	4	1250	1420
97	9	7043	6753
99	15	18522	18692
99	17	19717	19541
100	2	4094	1980
103	1	48	299
103	6	4924	4373
104	5	6142	6735
105	7	6098	6517

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
106	1	1	363
106	10	9832	10212
108	1	2	268
111	3	3417	3788
111	4	3809	4606
115	10	10854	10438
116	3	2873	2121
118	2	2274	1357
122	4	2698	2333
122	10	5858	6199
122	12	6301	7416
124	2	346	690
128	4	2544	3368
129	1	689	102
129	2	1011	724
129	8	6454	6056
129	9	6540	6277
129	12	7809	7621
131	3	1433	756
131	10	5972	5673
134	11	11838	11209
135	2	625	1140
136	4	2913	3830
137	2	325	134
139	12	14027	14521
139	13	14840	14532
139	14	15363	14875

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
140	20	19822	20838
142	1	1	285
146	3	760	479
146	4	1149	778
146	7	3604	2885
146	13	8223	9401
146	14	9399	10676
146	15	10052	9750
147	7	7488	7276
147	9	8913	8647
148	7	5298	4765
149	1	2	1936
149	3	2557	2880
149	9	6258	6070
150	2	1355	579
150	3	2556	1909
153	3	2061	2642
154	3	1953	1741
155	2	2181	1411
156	8	4550	4311
157	1	37	294
159	2	631	780
159	4	1384	1722
159	7	3271	4017
161	2	1332	1018
165	3	5535	4945
166	6	5406	4972

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
167	9	6075	6395
169	5	2828	3205
170	7	6485	6243
170	8	6964	6362
170	9	7303	6962
170	11	8790	7906
171	9	7150	7476
172	5	2298	1948
173	4	2913	2677
175	2	659	835
175	3	893	1789
176	2	1487	546
176	3	2200	1466
177	9	4686	4925
177	10	4923	5177
177	11	5111	5347
177	13	7396	8703
178	6	3452	3724
181	5	1853	2473
182	2	2112	1102
182	3	2617	2006
183	2	2126	2320
185	5	4683	4219
185	6	4846	4634
187	4	2940	3557
188	4	3686	4363
188	5	4183	4821

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
188	6	5882	6493
189	5	3143	2844
189	9	5956	5564
191	1	618	4
191	11	10357	10001
192	3	2861	2268
192	4	3081	2878
192	7	6800	5331
193	3	997	839
194	4	2315	2127
195	5	6249	4543
195	6	6620	6231
196	2	1553	1849
197	1	1	861
198	9	6844	6644
200	5	5329	5769
200	6	5993	6595
204	5	3914	3276
205	2	447	1709
209	4	2038	2460
209	5	2458	2682
210	10	7370	8230
210	13	9029	10441
210	14	10439	10705
214	5	2581	2330
214	9	5065	5277
214	11	5996	5754

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
217	2	541	194
218	2	914	1432
218	3	1430	1972
218	6	3639	3821
219	1	458	39
220	1	869	600
223	4	2617	1964
227	1	1	510
234	4	1539	1312
234	6	2116	1838
235	1	52	312
235	2	310	687
238	1	660	64
246	1	1	270
248	1	3	362
248	2	443	1222
254	3	2789	792
258	2	1179	1616
260	3	1770	2123
263	1	653	177
263	4	2244	1900
263	5	3569	2973
266	1	1	342
266	2	177	1022
270	2	1124	1681
272	1	857	186
275	2	1684	2295

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
278	1	2	406
282	1	714	391
282	4	1463	1134
287	2	1119	826
288	1	540	4
289	1	684	4
291	5	1589	1858
293	2	2539	2925
294	1	21	608
296	2	494	700
296	3	670	843
302	1	261	530
309	3	559	350
310	2	249	1889
316	2	2087	1818
317	2	1048	584
318	2	313	777
319	3	477	133
327	2	912	607
331	1	1	549
333	1	2	535
333	2	465	82
333	3	127	342
341	1	1	705
345	2	895	701
346	2	750	199
349	1	1	198

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
350	2	81	413
355	1	44	973
358	2	636	448
360	2	948	628
364	2	1639	1265
378	1	345	1004
379	2	683	510
381	1	109	693
385	1	150	4
385	2	269	30

(1) GENERAL INFORMATION:

(i) APPLICANT: Charles Kunsch
Gil H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
Brian A. Dougherty

(ii) TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

(iii) NUMBER OF SEQUENCES: 391

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Human Genome Sciences, Inc.

(B) STREET: 9410 Key West Avenue

(C) CITY: Rockville

(D) STATE: Maryland

(E) COUNTRY: USA

(F) ZIP: 20850

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

(B) COMPUTER: HP Vectra 486/33

(C) OPERATING SYSTEM: MSDOS version 6.2

(D) SOFTWARE: ASCII Text

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Brookes, A. Anders

(B) REGISTRATION NUMBER: 36,373

(C) REFERENCE/DOCKET NUMBER: PB340P1

(vi) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (301) 309-8504

(B) TELEFAX: (301) 309-8512

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5625 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CCAAGCAAAA CCAGCTACAG CTAAAGGAAC TTACGTAACA AACTTGACTA TCACAACTAC 60

TCAAGGTGTT GGTATCAAAG TTGACGTAAA CTCACTTTAA TCAGTAGTTA AAGTAATGTA 120

AAAAAGTTGA AGACGCTATG TCTCAACTTT TTTTGATGTA CGACGGGCAT GTTGTATAGT 180

AGATGTGTAC TATTCTAGTT TCAATCTACT ATAGTAGCTC AGAAGTCGGT ACTTAAACGT 240

GCTATATCAA AACCAGTCCT TGAAAAACGT GGACTGGTTT CGTGTTTGGA TTATTACCTT 300

GAACGACATG CGTTAAAAGT TAGTTGAACC GCCGTATGCC GAACGGACGT ACGGTGGTGT 360

GAGAGGGGCT AGAGATTATC CCCTACTCGA TTTCGAAATC TAGTGGAATG AATCTGGAAT 420

AGTCCATCGA GCTTTCTAAT ACTCTTCGAA AATCTCTTCA AACCACGTCA ACGTCGCCTT 480

GCCGTGCGTA TGGTTACTGA CTTCGTCAGT TCTATCCACA ACCTCAAAAC AGTGTTTTGA 540

GCTGACTACG TCAGTTCCAT CTACAACCTC AAAACAGTGT TTTGAGCAAC CTGCGGCTAG 600

TTTCCTAGTT TGCTCTTTGG TTTTCATTGA GTATAACACA TTGTTAGAAG TTGGTTTAAA 660

TTTCCTAATC AGTTTGTTCA CATTTACCTT CGATATATTA TATCCCATAG TTAAGGTTGG 720

TCATACAGAT GATTATAGTC ATGGAGCCGT AAAACTTAGT GTTTCTTTAG TTGACAAAGA 780

TGCCATGAAA AAAATATTTG TAACTGTAAT AGGATATTTT GAAATAAATA TAGATGAAAA 840

TATCACCGAT ATTCTATACG TAAATGGTAC TGCTATTCTT TATCTTTATT TACGTTCAAT 900

TGTTTCAATA GTTTCGGCAA TTGATAGCAG TGAAGCAATG TTGCTACCTA TCATTAATGT 960

TTTAGAGTTA CTAGATAAAT CTCAACCTTT TGAAGAAGAA TAATTTATTA GCTCACTAAA 1020

TTGAGGGTAA GGAAAAGTAA AAGCAGTAAG AAAAATGTCT TGCATTATAC AGCAACCTTT 1080

TGGGAATGAG TGGATGGATT GAATAAAATT TGATTAAGAG TGGATGATTT ATCTGTAGAT 1140

TATTATTGGA CAGTTAGTCT TGAAGTAGTC TAAGAATTAG GTTATAATCA GTAGAAGCCT 1200

TGCTAATAAT GAGGAGGTTA GTTTATGTAT AGTAGACTGA ATCTAAAATA GTACGAAACA 1260

ATTGCTAAAA CATTTATAGA AATTAATTTT ACTTTCCCAA TCGATTTGTT CTCATCTTAT 1320

TTCAATCCGC TATATATTAT GGTATCGAAT CTTCATCAGA ATGATAAAAT TAATCAATTG 1380

ATATCTGATT ACAAACAGAA TATGAAAGCT TTTTATATCA CTATTGAAAA ATTTATACGA 1440

GATGATGAAA GCCTTAAGTG TTATTTTATA AAGGTTATTT CAAGTCGTTC CAAGGTAACA 1500

AGTCTAGATC AGATTGAAGC TGATAAAACG ATACAAAGAA AATATTCAAG TGAGCTAAAA 1560

AAATTTATTG GATTTTATAA TGAGATTATT TGTGAGGAAA ATAGTTTCCT ACATGTACGA 1620

AAGAGGTGGT CGAGTTGGTT TAGGTAGTCG ATGCGTGAGT TGATAATTCT CAGGGTATGG 1680

ACTTCTTTTT CATGAATGAG GTAAAAGAGC AGGTATTGTT TAGAGACAAT CATTCTGAGC 1740

ATATTTTCTG GATAGAGGGA GTATCCGATT TTATGATCAA AGTTAATACC GCCCTCTGGT 1800

GAGAAGATGA GTAGGTTGGT AATTTAAACT ATTAAACAGA ATTTTTGATT AAAAGTATTA 1860

TTTCATGAGA GAAATCCTAA TTTCACAATC CATAGGCAAA CGCTTGCATT TCGTTTTTTA 1920

TTGGACTATA ATAGGTTGGT ATAAAGCCTT CTGTAGTAAT AAAATGTAGA AGGTGTAGAA 1980

AGTAAGGATT TAGAATATTT GTAGTTAAAA ACACAATGTT GCTATTCCTT ACGATAGGGA 2040

GATAGATATG GCAATGATAG AAGTGGAACA TCTTCAGAAA AATTTTGTGA AGACTGTTAA 2100

GGAACCGGGC TTGAAGGGGG CTTTGCGCTC CTTTATTCAT CCTGAAAAGC AGACCTTTGA 2160

AGCGGTCAAG GATTTGACCT TTGAGGTTCC AAAAGGGCAG ATTTTAGGAT TTATCGGGGC 2220

AAATGGTGCT GGGAAGTCGA CAACCATTAA AATGCTGACA GGAATTTTGA AACCAACATC 2280

TGGTTTTTGT CGGATTAACG GCAAGATTCC CCAGGACAAT CGGCAAGATT ATGTCAAAGA 2340

TATTGGCGTA GTCTTTGGAC AACGCACCCA GCTATGGTGG GATTTGGCTC TGCAAGAGAC 2400

CTACACTGTC TTAAAAGAGA TTTATGATGT GCCAGACTCG CTCTTTCATA AGCGTATGGA 2460

CTTTTTGAAT GAAGTCTTGG ATTTGAAGGA CTTTATCAAG GATCCCGTGC GGACTCTTTC 2520

ACTGGGACAA CGGATGCGGG CGGATATTGC GGCCTCCTTG CTCCACAATC CCAAGGTTCT 2580

TTTTTTAGAT GAGCCGACCA TTGGTTTGGA CGTTTCGGTT AAGGATAATA TTCGTCGGGC 2640

AATTACTCAG ATCAATCAAG AGGAAGAAAC TACCATTCTT TTGACCACTC ACGATTTGAG 2700

TGATATTGAG CAACTTTGTG ATCGGATTTT CATGATTGAC AAGGGGCAAG AGATTTTTGA 2760

TGGAACGGTG AGCCAACTCA AGGAGACCTT TGGTAAGATG AAGACTCTCT CTTTTGAACT 2820

GCTACCAGGT CAAAGTCATC TCGTCTCTCA CTATGACGGT CTGTCTGATA TGACCATTGA 2880

TAGACAAGGA AACAGCCTCA ACATTGAATT TGATAGTTCT CGCTACCAGT CAGCTGACAT 2940

TATCAAGCAA ACCCTGTCTG ATTTTGAAAT CCGCGATTTG AAGATGGTGG ATACGGATAT 3000

TGAGGATATT ATCCGTCGCT TCTACCGAAA GGAGCTCTAG GATGATCAAA TTGTGGAGAC 3060

GTTATAAACC CTTTATCAAT GCAGGGGTTC AGGAGTTGAT TACTTACCGA GTCAACTTTA 3120

TTCTCTATCG GATTGGCGAT GTCATGGGGG CTTTTGTGGC CTTTTATCTC TGGAAGGCTG 3180

TCTTTGATTC TTCGCAAGAG TCTTTGATTC AGGGCTTCAG TATGGCGGAT ATCACCCTCT 3240

ACATCATCAT GAGTTTTGTG ACCAATCTTC TGACTAGATC CGATTCGTCC TTTATGATTG 3300

GGGAGGAGGT CAAGGATGGC TCCATTATCA TGCGTTTGTT GCGACCAGTG CATTTTGCGG 3360

CCTCCTATCT TTTCACCGAG CTTGGTTCCA AGTGGTTGAT TTTTATCAGC GTTGGCCTTC 3420

CATTTTTAAG TGTCATTGTC TTGATGAAAA TCATATCGGG TCAAGGTATT GTAGAGGTGC 3480

TAGGATTAAC TGTCATTTAT CTTTTTAGCT TAACGCTCGC CTATCTGATT AACTTTTTCT 3540

TTAATATTTG CTTTGGATTT TCAGCCTTTG TGTTTAAAAA TCTTTGGGGT TCCAACCTAC 3600

TTAAGACTTC CATAGTGGCT TTTATGTCGG GGAGTTTGAT TCCCTTGGCA TTTTTTCCAA 3660

AGGTTGTTTC AGATATTCTC TCCTTTTTGC CTTTTTCATC CTTGATTTAT ACTCCAGTTA 3720

TGATCATTGT TGGAAAATAC GATGCCAGTC AGATTCTTCA GGCACTCCTT TTGCAGTTCT 3780

TCTGGCTCTT AGTGATGGTG GGATTGTCTC AGTTAATTTG GAAACGGGTC CAGTCCTTTA 3840

TCACCATTCA AGGAGGTTAG TATGAAAAAA TATCAACGAA TGCATCTGAT TTTTATCAGA 3900

CAATACATCA AACAAATCAT GGAATATAAG GTAGATTTTG TGGTTGGTGT CTTGGGAGTC 3960

TTTCTGACTC AAGGCTTGAA TCTCTTGTTT CTCAATGTCA TCTTTCAACA TATTCCATTC 4020

CTAGAAGGCT GGACCTTTCA AGAGATAGCT TTCATTTATG GATTTTCCTT GATTCCCAAG 4080

GGAATGGACC ATCTCTTTTT TGACAATCTC TGGGCACTAG GGCAACGCCT AGTCCGAAAA 4140

GGGGAGTTTG ACAAGTATCT GACTCGTCCC ATCAATCCTC TCTTTCACAT CCTAGTTGAA 4200

ACCTTTCAGA TTGATGCCTT GGGTGAACTC TTAGTCGGTG GTATTTTATT GGGAACAACA 4260

GTGACCAGCA TTGTTTGGAC TCTTCCAAAA TTCCTGCTTT TCCTAGTTTG TATTCCTTTT 4320

GCGACCTTGA TTTATACTTC TCTTAAAATC GCAACAGCCA GTATCGCCTT TTGGACTAAG 4380

CAGTCAGGCG CCATGATTTA CATCTTCTAT ATGTTCAATG ACTTTGCTAA GTATCCGATT 4440

TCTATTTACA ATTCTCTTCT TCGTTGGTTG ATTAGCTTTA TCGTGCCTTT CGCCTTTACA 4500

GCCTACTATC CAGCTAGCTA TTTCTTACAG GAAAAGGATG TGTTCTTTAA CGTAGGAGGT 4560

TTGATGTTGA TTTCTCTGGT TTTCTTTGTT ATTTCCCTTA AACTTTGGGA TAAGGGCTTA 4620

GATTCCTACG AAAGTGCGGG TTCGTAAAAG CTAAAGTAAG ACTAAAATCA AGAAAGAAAC 4680

TTATGATGTT TGTAATTGAA GAAGTCAAGG ATGAAAATCA AAAAAAGGCA GTTGTCGCTG 4740

AGGTTTTGAA GGATTTGCCA GAATGGTTTG GAATCCCAGA AAGCACACAA GCCTATATAG 4800

AAGGAACCAC GACACTGCAA GTTTGGACCG CCTATCAGGA GAGTGATTTG ACTAGATTTG 4860

TAAGCTTATC CTATTCGAGT GAAGATTGTG CAGAGATTGA TTGTCTCGGC GTAAAAAAGC 4920

TTATCAAGGT AGAAAAATTG GGAGCCAATT GCTTGCTACT TTAGAGAGTG AAGCTCGTAA 4980

AAAAGTTGGT TATCTGCAGG TCAAAACAGT GGCAGAAGGT TCTAATAAAG ATTATGATCG 5040

AACAAATGAC TTTTATCGAG GTCTTGGCTT TAAAAAGTTA GAGATTTTTC CTCAACTATG 5100

GAATCCGCAA AATCCTTGTC AGATTTTGAT TAAAAAGCTT GAATAATATT ACTTGACATC 5160

TATTCTCAGA GTGCTATACT GTAAGTGTAA TCGCCGATTT AGCTTAGTTG GTAGAGCAAG 5220

GCACTCGTAA AGCCTAGGTT ATAGGTAGAT AAACGACTGA GGATTTGAAA AAATAGATAG 5280

GTAGAAGATA ACCGTTAAGC CTTACTCTTA GCGGTTATTT ATATTGTTTA ATAGCGCTAA 5340

TATTTTATCA ATTATGCCTG TTTTCGTGTT TCTGGTAGTT GTTCAAGTTT ATTGCTACTA 5400

TTTTTGATGG TATGAATGTG CTTATAATGT ATCCCGGTTA ACGAAAGTTT TGGACTTATA 5460

CTCTTCGAAA ATCTCTTCAA ACCACGTCAA CGTCGCCTTG CCGTGCGTAT GGTTATGACT 5520

TCGTCAGTTC TATCCACAAC CTCAAAACAG TGTTTTGAGT GACTACGTCA GTTCCATCTA 5580

CAACCTCAAA ACACTGTTTT GCCCAATCTG CGGCTAGTTT CCTAG 5625

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7571 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CTCTCCAGCT TTCCTTGCGA GTTGGCCATG TTGTGTCTTT AAGAAGTCTA AAAATATCTC 60

CAATAAAACG CATCGCTCTC TCCTATCTCG TTTCTCTGTG TGTAGTGTAC TTGCCACAAT 120

GCTTACAAAA TTTATTTACT TCTAGTCGTG TAGGCTTGAG GTTTCCGCTG ATCTTGATTG 180

AATAGTTTCT CGAACCACAA ACCGCACAAG CTAGGCTTGC TTTTTTTAGT GCCATAACGC 240

CTCCATCTTA TCCATTATAA CAAGAAAGCT AGGCTTTGAC AAGCATCTTA GCGAAATAGA 300

TTGACTATCG AATCCCATAT TGTTTGAGCC TTTTCCTTAA TCTTCGCATC TGAGATAGCC 360

CGGCTAGCCT CATCTACTAG ACTTTGCGCA CGCCCTCGAA TATCAGACAA ATTATCATCT 420

GTCTGGCTAT TATCATTGGT TTGTACTTGT CTTTTTGTAT TGGCTGGTGC AATTCCATTT 480

TGCTTATAAG CATTTTCAAC CGTAAAGGTA CTTCCTGGCG TATAAGGTAA AATGGTATTG 540

GCAATGTTTC TAAAGACATG AGCTGCACCG TTTGAAGTAG AGCCAGCTAG ATAGTGGTTT 600

TCATCAGTGG TCGGAAAGCC AAGCCAGTGG CTAATCACTA CATCCGGAGT ATAACCAATT 660

ACCCACTGGT CACTTGTGTA CTCCGGATTG AAAACTGCTT CAGTTGTTCC AGTTTTCCCT 720

GCCATGACAT AGTCTGCAGG CGATGAACTA ATACCGGTAC CGTTGGTGAA AGTCCCCAAC 780

ATCATACTGG TCATCTTGTC AGCTACAGAC TTATCAATCA CCCGTTTTTG TGAATTTTTA 840

TGACTCGCAA TAACTTGTCC ACTAGCATTT TCAATTCTAC TAATAAAATG AGCTTCAGGC 900

ATTAAACCTT CATTTGCAAA GGCGGCGTAT GCTTGAGCCA TTTGAAGAGG GTTGGTTTCA 960

ACACCGCTTC CCAAGGCGAC ACCAAGAACA CGGTCGACCT TTTCCATGTT GAGTCCGAAT 1020

TTTTCGCCTG CCTCAAAAGC CTTGTCGACA CCCAAATCAT TAACAGTGGC AACAGCAGGT 1080

AGATTAAGCG ATTCTGCCAA GGCTTGATAC ATAGGAACTT CTCGACTCGT TTTGATCCCT 1140

GCATAGTTAT CAACCTTATA GCTGTCATAC TGCATGGTAT GGTTATCCAA CTGCTTATTC 1200

AAAGCCCAGC TTGCTTCAAC TGCTGGCGTA TAAACAACTA AAGGCTTAAT TGTAGAACCA 1260

GGACTACGCT TTGATTGGGT TGCATAGTTG AAATTCCGGA ATCCAGTTTT ATCATTGTCA 1320

GCAACTTGAC CGACAACTCC ACGAACTCCC CCTGTTTTCG GTTCGAGGGC TACACTTCCT 1380

GATTGAGCAA ACGTTCCATC CTCTGCCCTC GGAAATAGCG ATGTGTTTTC ATAAACAATC 1440

TGCATATTTG CTTGGTAGTT TTGGTCCAGC TCTGTGTAAA TGCGGTAGCC ATTATTGACA 1500

ATCTCTTCCT CTGTTAGATT ATACTTGGAA ACAGCTTCAT TAACCACCGC ATCAAAATAA 1560

GAGGGGTAAC GGTAATCTGA GATTTTTCCT TCATACTTAT CGTGCAATTG CGAAGTCATA 1620

TCAACTTCAG CAGCTTTGGT TTCTTGGTTT TTATCAATAT ATCCTGCTGC AACCATATTC 1680

TGCAAGACAG TATCGCGCCG ATTAGTAGAA TCTTCTACGG AATTCAAGGG ATTATACAGT 1740

TCCGGCCCCT TGAGCATCCC TGCCAGAGTC GCAGCTTGAT CCAGACTCAC TTCTGATGCA 1800

GAAACTCCAA AGTATTTCTT ACTCGCATCT TCTACACCCC ACACACCATT TCCAAAATAA 1860

GCGTTGTTAA GGTACATGGT TAGAATTTGC TCCTTACTAT ATTTTTTGCT TAATTCTAAG 1920

GCAAGGAAAA ATTCTTTCGC TTTTCTCTCA ACAGTTTGAT CCTGCGATAA ATAGGCGTTT 1980

TTAGCCAGCT GTTGGGTAAT GGTAGAGCCA CCACCTGAAC GTCCAGCAGT GACAATAGCC 2040

AAGAAAAAAC GGCCATAGTT AATCCCGTCA TTTTTATAGA AAGAACGGTC TTCTGTCGCA 2100

ATAACAGCAT TCTGCAAGTT TTTACTGATG TCAGTCAGCT CAACATAGGT TCCCTTTTGA 2160

CCAGACAAGG CACCAGCCTC TTTTTCTTCA CGGTCAAAAA TAAGAGTCCG AGTTTTCAAG 2220

GCATTTTGCA AATCATTGAC ATTGGTCGAC TTGGCTACAG CAAACAAATA GATTCCAACT 2280

AGCAAGCCTG CACTCAAACC TAGTATAAGG ATAATCTTTG TTAGATGATA ACGACGCCAG 2340

AATTTTCGAA TCGGACCTAC TTGGGCTAAT TTTTTTCGAT CACTACGAGA GCGACGTAAG 2400

ATAGTAGAAT CAGAGTCCTC TAGTTCACTT GTTTCTTTTT TAAAAAGAGA AAGAAATTTC 2460

TCAAATAATT TATCTAATTT CATGCGTTTA TTTTATCATC TTCATCATAG GAAGACAAGA 2520

ATTTAGCTAT TTCCTATCCA AATAGGGCTT TTTTTGTTAC AATATCTGTA TGCAATTCAC 2580

ATTTACATTA CCCGCCTCTC TACCTCAAAT GACAGTAAAG CAATTACTTG AGGAACAACT 2640

CCTCATCCCT AGAAAAATCC GTCATTTTTT GAGAATCAAG AAACATATTT TGATAAATCA 2700

AGAAGAAGTC CACTGGAAGG AAATCGTAAA TCCTGGAGAT GTTTGCCAGT TGACTTTTGA 2760

CGAGGAAGAT TATTCCCAAA AGACGATCCC TTGGGGCAAC CCAGACTTAG TGCAGGAAGT 2820

TTATCAAGAT CAACACTTGA TTATTGTAAA CAAACCAGAG GGGATGAAAA CGCATGGTAA 2880

TCAACCAAAC GAAATTGCCC TTCTTAACCA TGTCAGTACC TATGTTGGCC AAACCTGCTA 2940

TGTCGTTCAT CGTCTGGACA TGGAAACCAG TGGCTTAGTT CTCTTTGCCA AAAATCCTTT 3000

TATCCTGCCC ATTCTCAATC GCTTATTGGA GAAAAAAGAG ATTTCTAGAG AATATTGGGC 3060

TCTAGTTGAT GGAAATATCA ACAGAAAAGA ACTTGTTTTC AGAGACAAAA TTGGACGTGA 3120

TCGCCATGAT CGTAGAAAAA GAATAGTTGA TGCAAAAAAT GGGCAATATG CTGAAACGCA 3180

TGTAAGCAGA TTAAAGCAAT TCTCAAACAA GACTTCCTTG GCTCATTGCA AGCTAAAGAC 3240

AGGGCGAACC CATCAGATTC GTGTGCACCT TTCGCATCAT AATCTTCCTA TCCTGGGAGA 3300

CCCTCTCTAT AATAGTAAAT CAAAGACAAG CCGGCTTATG CTTCATGCCT TCCGACTTTC 3360

CTTTACCCAC CCACTTACTT TAGAGAAGCT AACTTTCACT ACCCTTTCAA ATACATTTGA 3420

AAAAGAATTA AAAAAGAATG GATGATCGTG TCATCCATTT TTCCATATAA AAAAGCAAGA 3480

CCACAAAGCC TTGCTTTCTA TCAACTCAAG AATTATTTAG CAATTTTTGC GAAGTATTCA 3540

AGAGTACGAA CAAGTTGTGC AGTGTATGAC ATTTCGTTGT CGTACCATGA TACAACTTTA 3600

ACCAATTGTT TACCGTCAAC GTCAAGAACT TTAGTTTGAG TTGCGTCAAA CAATGAACCG 3660

TAAGACATAC CTACGATATC TGAAGATACG ATTGGATCTT CTGTGTAACC GTATGATTCG 3720

TTTGAAGCTG CTTTCATAGC TGCGTTCACT TCATCAACAG TAACGTTCTT TTCAAGAACT 3780

GCTACCAATT CAGTAACTGA TCCAGTTGGA GTTGGAACGC GTTGTGCAGA TCCGTCAAGT 3840

TTACCATTCA ATTCTGGGAT TACAAGACCG ATAGCTTTTG CAGCACCAGT TGAGTTAGGA 3900

ACGATGTTTG CAGCACCAGC GCGAGCACGG CGAAGGTCAC CACCACGGTG TGGTCCGTCA 3960

AGGATCATTT GGTCACCAGT GTAAGCGTGG ATAGTAGTCA TCAATCCTTC AACAACACCA 4020

AAGTTGTCTT GAAGAGCTTT AGCCATTGGA GCCAAGCAGT TTGTAGTACA TGAAGCACCT 4080

GAGATAACTG TTTCAGTACC GTCAAGAACG TCGTGGTTAG TGTTGAATAC AACTGTTTTA 4140

ACGTCGTTTC CACCAGGAGC AGTGATAACA ACTTTTTTAG CTCCACCTTT AAGGTGTTTT 4200

TCAGCTGCTT CTTTCTTAGC AAAGAAACCA GTAGCTTCAA GAACGATTTC TACACCGTCA 4260

GTAGCCCAGT CGATTTGTTC TGGATCACGT TCAGCAGAAA CTTTGACTGAA TTTACCGTTA 4320

ACTTCAAATC CACCTTCTTT AACTTCAACA GTACCGTCGA AACGACCTTG AGTTGTGTCG 4380

TATTTCAACA AGTGTGCAAG CATAACTGGA TCTGTAAGGT CGTTGATGCG TGTAACTTCA 4440

ACACCTTCTA CGTTTTGGAT ACGACGGAAA GCAAGACGAC CGATACGTCC GAAACCGTTA 4500

ATACCAACTT TAACTACCAT TAGTGATTTC CTCCTTATGA AAATCATGAA ATTTTTATTG 4560

TGAAAAGAGT AACTTGAATC ACTACAAATC ACCTTTCAAC AAACCTATTA TACAACTATT 4620

TGAGTTGAAT TGCAAGTATG GCCATTGTTT TTCTATGTTA GTTTCTTTTT AAGACTGTAA 4680

ACCAAGGAAT CCCTTACTAT TCATAGCATA ACGATTCTAT AGGATCCATT TTACTAATCT 4740

TACGCGCCGG GAAGTAGGCT GAGACATAAC CAAGTAATAG AGCGAAAACT AGAGTTCCTA 4800

AAACAGATAA AAGATTTAAT TTAAAAACCT TAGTGATGGA TGGGTAAAAG TGACTTACAA 4860

TCGCATTCGC CAAACTTCCC ACCCCTTGTG CAACCAAAAA TGCCAGCAGC AAGGCGATGC 4920

CTACAATCCA GATAGCCTCG TAAATAAAAA TTCCTTTGAC ATCACGATTC TGATAACCAA 4980

CTGCTTTCAT GACACCTATT TCCTTGGAAC GTTGCATGAT ATTGATGTAA ATAATGATAC 5040

CAATCATAAC CGCTGCTACC ACAATAGCTT GTGATGAAAG CACAATCAAT AATCCCTGAA 5100

TAACACGAAT AAAGGTAATC ACAATATCAA GAACTCTCTG TTGAGAAAGC ACAGTATACT 5160

TCTTATTTTT CTGTAATTCT TCTGTTACTA CTTTTGTCTG TGATGGATCT TTGAGTTCCA 5220

AGATAAAATA AGATACAGCT TTCGTAAATC CAGCCTCTTT CAAAATCGTT TCCATTTGAT 5280

GAGACAGCAT GAAACTGTTG CTGTCCTCCA TGTCATCTTC ATCATTGATT ACACGTACAA 5340

TCTTCGTTTG AAATTGAGCA ATCTTACTAG TTTCGGCAGC ACTTTCTACA ATGCTGGCTG 5400

AGACTGATTT GCCAATAAGA TCATTAGCTG TCAAATTTTT TCCTGTCTGT TCATTCCAAT 5460

TTTTTAGTAA ACTGCTTGGA ATCGTTAATC CCTGTTCATT TGTATCAGTA TAGAGGGATC 5520

CAGCCAACAC TTTGTCCGTC TCATTATTAC TAACAGAGAT ACTTGTATCA TCATAAAGAC 5580

TCACTACTTG AGCATAAGAA GGCATCGTTT GACTCAGATC CATTTCTTGC CCATCTATAG 5640

TAATATTTGA CATGTTCATC CCAAAAGGAC TCTCCAAATA TTTAATAGCT TCTTTCCCAA 5700

CTGTATCCGT GATATATAGT CAATTGAAAC AAGAGCAGGA TAAAAAAGCC TCGTAAAAGG 5760

TATTGCAACT TGGTAATACC TTTTTGAGGT GCTTTTTGAT ATGAGCCCAT GTTTTCTCAA 5820

TAGGATTGTA CTCAGGCGAG TAGGGAGGAA GAGGTAAAAG TTTATGCCCA AACTCTTCGC 5880

ATAAAAGTTC TAGCTTCCCC ATTCTATGGA ATCTTACATT ATCCATAATA ATAACCGATG 5940

GTGTGTTTAA TGTTGGTAAG AGAAAATTCT GAAACCAAGC TTCAAAAAAG TCGCTCGTCA 6000

TCGTCTCTTC GTAAGTCATT GGAGCGATTA ATTCACCATT TGTTAGACCT GCAACCAAAG 6060

AAATCCTCTG ATATCTTCTT CCAGATACTT TGCCTCTTAT TAATTGACCT TTTAATGAGC 6120

GACCATATTC TCGATAAAAA TAAGTATCGA ATCCTGTTTC GTCAATCTAA ACAGGTGCTA 6180

GGTGCTTTAA ACTATTAAAA TTCTTAAGAA ATAAGGCTAC TTTTTCTGGG TCTTGTTCAT 6240

AGTAGGTGTG GTTCTTTTTT CGAGTGTAGC CCATAGCTTT GAGCGTATAG TGGATGGTAG 6300

TTGGATGACA GCCAAATTCA GAAGCTATTT CAGTCAAATA AGCGTCTGGA TTGTCAGTAA 6360

GATAGTTTTT AAGTCTATCT CTATCAACCT TTCTTGGTTT TATTCCTTTT ACTTGGTGGT 6420

TTAGCTCTCC TGTTTTCTCT TTTAGCTTTA ACCAGCCATA AATGGTATTA CGTGAGATTT 6480

GGAAAACGTG TGATGCTTCT GTTATACTAC CTGTTCGCTC ACAATAAGAG AGAACTTTTT 6540

TACGAAAATC TATTGAATAT GCCATAAAAA GATTATACCA CATTGTGTAC TATTTTTGGT 6600

TCATTTTACT ATATTTGAAG AGGCGTTTAA ACTATCTGAC ATAAAACTCG TTCTAGAGGA 6660

AAGACATCCT TTAAAAAGTT AGTTTATTTT ACAACTTAGA CATCAAGGTA GGTTAACCCC 6720

TTCATGGAAA AATCAAGACT CTTAGCACTA TGGGTTAAAC TACCACTGGA GACGTAATCA 6780

ATCGCTAAAC CACGAAAACG GCTAATAGTG GTCATATCAA TATTTCCAGA ACATTCAATC 6840

CGAGAACGTC CTGCAATTAG GGTAATGGCC TGTTCAATCT GTTCCAATGA CATATTATCC 6900

AACATGATAA TATCAGCACC CGCCGCCGCA GCTTCTTCGG CAGCAGCAAG GCTTTCCACT 6960

TCCACCTCGA CCATTTTCAC AAAAGGGGCA TAGGCACGCG CTTGAGCAAT TGCCTTTTGA 7020

ACACTACCTA CTGCCGCAAT GTGATTGTCT TTTAGCAGGA TAGCATCTGA TAAATTAAAG 7080

CGATGATTAT AGCCACCGCC AACTCTCACG GCATATTTCT CAAAAAGACG TAAATTAGGA 7140

GTAGTTTTTC GAGTATCAAA TACCTTAATG CAATCATCGC CTAAGGCTTC TACATAAGCA 7200

GCTGTCATCG AAGCAATCCC TGATAAATGT TGTAAAAAAT TCAAGGCAAC GCGTTCACAT 7260

GTTAAGAGAC TTCTCACCGA GCCTATGATT TCTAAAACCA AATCGCCACT AGTCAAACGA 7320

TCCCCATCCT TAAATTGATG AGGATTCTGG AAGGTCACCT CGGCATCAAA TAGGGTAAAA 7380

ACCCTTTGAA AAACGGTTAG CCCCGCTAAA ACACCAGCTT CCTTGGCAAA AAGCGACACC 7440

TTGGCTTGGC CATGATGATC AAAAATGGCA TTGGTACTGT AATCTTCGGA ATGAACATCT 7500

TCTCGCAAGG CTGCTTTCAA TGTATCATCT ATTTGAAAAG GGGTTAAATC AGTTGAAATG 7560

ATTGACATCA C 7571

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26385 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TTTGCTAGTG GCTTAAATTC TTCAGGAAAA TCAGGCGTAT CTAAAAGTCG TGTCGTTTTT 60

GTTTCATCTA TATAAAGACT TCCTGCTCCC CCTACAACTA GAAAACGTGT CTGTGTTCCA 120

GCAAGAAGCT GATTAAATAG TTCGATTGAT TTGCTGTGGA GCGGTAGCGT ATCTGGTGTA 180

TAAGCACCAA ACGCTGAAAT AACAGCATCA AATCCAGTAA GATCATCTTT TGTCAACTCA 240

AATAAATCTT TTTTAATAAT AGACTCAGCT TGACTTTTGT TTTCAGAACG AACAATAGCC 300

GTTACTTCAT GTCCTCGTTT GACTGCTTCT TCAACAATTG CTTTCCCCGC TTGTCCATTT 360

GCTGCAATAA CTGCTAGTTT CATTTTTTAT ACCTCTCTTG TTGTAATTAT TTTAGTTACA 420

GAAATTGTGA CACTCTTAAT AATCAATGTC AATAGTCTTG CTTAATTATT ATCAAAATAT 480

TTCTACCAAG AAAACTAACC ATGATTCTAG TGAAAAAAAA TCTTCTTTGT CAACAAATTT 540

ACTTTCTTGT TTTAAACATG CTATAATAAT CATAGCAAGA GATCTAAGTT GTCTGTTTTT 600

TTAAAACGAG GTGATTATCA TGCGTAGATT CTATTCCCAT CTCCCCTACT ATCTGGTCAT 660

ATTATTCTTT TATTGGCCAC TTTATGAGTT GTTCTTACTA GTTGTTTCTG ACCCCCTTAC 720

ACTCAAGGGA CTCTATATAA ACAATCTTCT CTTCTTTACA CCTCTGGTAA TCTTGATTGT 780

ATCGTTACTC TATAGCTACC GTTTCCGTTT CTCACTTTGA TGGTTAGTTG GTAACGGACT 840

GCTCTTTTAC TTTACTATCA TAACCTTTGG TGAGTTTATA CTAATTTACT TGCTAATCTA 900

TGAAACAGTT GCTCTGGTCG GCATGGATTC TGGTATTAGC ATCAAGCATA TTCTACAAAA 960

AATGAAAAAC AAAAAACTTT CACAAAATCC TTGAAAAATC TCACAATCAT GCTATAATAA 1020

TCCATAGAGA CAAGTCACTT AGTCCCTTTC TACTAGAGAG TGCGTGGTTG CTGGAAACGC 1080

ATAGGAAGTC TAAACTGATA CTACTCTTGA GTTTTTTATG AAAACATAAA ACGGTGGCCA 1140

CGTTAGAGCC GATCAGAGGT GTCCCTCTCT TTTGAGGTAC ATAAATGAAG GTGGAACCAC 1200

GTTGCGACGT CCTTTCGAGG ATGTCGCATT TTTTTATTAG GATACTAATT ATGGAGTTGC 1260

AAGAATTAGT GGAGCGCAGT TGGGCAATCC GACAAGCTTA TCACGAACTG GAAGTTAAGC 1320

ATCATGATTC CAAGTGGACG GTAGAAGAAG ACCTCTTGGC TTTATCTAAT GATATTGGAA 1380

ATTTCCAACG ACTGGTGATG ACAAAGCAAG GACGCTACTA TGATGAAACA CCCTACACAC 1440

TGGAACAAAA ACTTTCAGAA AATATCTGGT GGCTATTAGA ACTTTCTCAA CGTTTGGATA 1500

TAGACATTCT GACGGAAATG GAAAACTTCC TCTCTGATAA AGAAAAGCAA TTGAACGTTA 1560

GGACTTGGAA GTAGTCTGCT GATAAAAAAT CAATGCTTAG AAACTATGAA ATAATAAAAA 1620

AGGAGAACAT CATGATTAAC ATTACTTTCC CAGATGGCGC TGTTCGTGAA TTCGAATCTG 1680

GCGTAACAAC TTTTGAAATT GCCCAATCTA TCAGCAATTC CCTAGCTAAA AAAGCCTTGG 1740

CTGGTAAATT CAACGGCAAA CTCATCGACA CTACTCGCGC TATCACTGAA GATGGAAGCA 1800

TCGAAATTGT GACACCTGAT CACGAAGATG CCCTTCCAAT CTTGCGTCAC TCAGCAGCTC 1860

ACTTGTTCGC CCAAGCAGCT CGTCGTCTTT TCCCAGACAT TCACTTGGGA GTTGGTCCAG 1920

CCATCGAAGA TGGTTTCTAC TACGATACTG ACAACACAGC TGGTCAAATC TCTAACGAAG 1980

ACCTTCCTCG TATCGAAGAA GAAATGCAAA AAATCGTCAA AGAAAACTTC CCATCTATTC 2040

GTGAAGAAGT GACTAAAGAC GAGGCACGTG AAATCTTCAA AAATGACCCT TACAAGTTGG 2100

AATTGATTGA AGAACACTCA GAAGACGAAG GCGGTTTGAC TATCTATCGT CAGGGTGAAT 2160

ATGTAGACCT CTGCCGTGGA CCTCACGTTC CATCAACAGG TCGTATCCAA ATCTTCCACC 2220

TTCTCCATGT AGCTGGTGCG TACTGGCGTG GAAACAGCGA CAACGCTATG ATGCAACGTA 2280

TCTACGGTAC AGCTTGGTTT GACAAGAAAG ACTTGAAAAA CTACCTTCAA ATGCGTGAAG 2340

AAGCTAAGGA ACGTGACCAC CGTAAACTTG GTAAAGAGCT TGACCTCTTT ATGATTTCAC 2400

AAGAAGTGGG ACAAGGTTTG CCATTCTGGT TGCCAAATGG TGCGACTATC CGTCGTGAAT 2460

TGGAACGCTA CATCGTAAAC AAAGAGTTGG TTTCTGGCTA CCAACACGTC TACACTCCAC 2520

CACTTGCTTC TGTTGAGCTT TACAAGACTT CTGGTCACTG GGATCATTAC CAAGAAGACA 2580

TGTTCCCAAC CATGGACATG GGTGACGGGG AAGAATTTGT CCTTCGTCCA ATGAACTGTC 2640

CGCACCACAT CCAAGTTTTC AAACACCATG TTCACTCTTA CCGTGAATTG CCAATCCGTA 2700

TCGCTGAAAT CGGTATGATG CACCGTTACG AAAAATCTGG TGCCCTCACT GGCCTTCAAC 2760

GTGTACGTGA AATGTCACTC AACGACGGTC ACCTATTCGT TACTCCAGAA CAAATCCAAG 2820

AAGAATTCCA ACGTGCCCTT CAGTTGATTA TCGATGTTTA TGAAGACTTC AACTTGACTG 2880

ACTACCGCTT CCGCCTCTCT CTTCGTGACC CTCAAGATAC TCATAAGTAC TTTGATAACG 2940

ATGAGATGTG GGAAAATGCC CAAACCATGC TTCGTGCAGC TCTTGATGAA ATGGGCGTGG 3000

ACTACTTTGA AGCCGAAGGT GAAGCAGCCT TCTACGGACC AAAATTGGAT ATCCAGATTA 3060

AAACTGCCCT TGGAAAAGAA GAAACCCTTT CTACTATCCA ACTTGATTTC TTGTTGCCAG 3120

AACGCTTCGA CCTCAAATAC ATCGGAGCTG ATGGCGAAGA TCACCGTCCA GTCATGATCC 3180

ACCGTGGGGT TATCTCAACT ATGGAACGCT TCACAGCTAT CTTGATTGAG AACTACAAGG 3240

GGGCCTTCCC AACATGGCTG GCACCACACC AAGTAACCCT CATCCCAGTA TCTAACGAAA 3300

AACACGTGGA CTACGCTTGG GAAGTGGCCA AGAAACTCCG TGACCGCGGT GTCCGTGCAG 3360

ACGTAGATGA GCGCAATGAA AAAATGCAGT TCAAGATCCG TGCTTCACAA ACCAGCAAGA 3420

TTCCTTACCA ATTAATTGTT GGAGACAAAG AAATGGAAGA CGAAACAGTC AACGTTCGTC 3480

GCTACGGCCA AAAAGAAACA CAAACTGTCT CAGTTGATAA TTTTGTTCAA GCTATCCTAG 3540

CTGATATCGC CAACAAATCA CGCGTTGAGA AATAAGAGTC TAGCATAAAA GCCTCCAATC 3600

TGGAGGCTTT TTCTCATCTA TTTTTACTCA AGGACTAAGT TCACTTGAGC AAACTGAATC 3660

CGCACTGTCG TTCCTTTTCC GACCTCAGAC TCGATACGAA TCTGGTGCCC CAGTTCTTCA 3720

GAAATTTTCT TAGATAGATA AAGGCCAAGT CCAGAGGACT GCTGGGTCAA ACGGCCATTG 3780

TATCCTGAAA AGCCACGTTC AAATACTCGG AGGACATCAC TGTTTTTTAT CCCGATTCCC 3840

GTATCTTTGA TACAAAGCTC TTGGTCATCC ATATAAATCT CCAGACCACC TTCCTTGGTG 3900

TACTTGAGAC TGTTTGAGAT GATTTGCTCA ATAACCACTA GCAGCCACTT TTTATCCGTC 3960

ACGATTTCTT TATCAAGGTC ATGTAGATTG ACATTTAAGC CTTTTTGAAT AAAGAAAAGA 4020

GCATATTTAC GAATTATTTC CTTGACCAAG TCCTCAATTT GAACCTGCTT TAAGACCAAA 4080

TCATCATGGA AACTTTCTAA ACGCAGGTAC TGTAAAACTA GGTTGGTATA GGAGTCGATT 4140

TTGAAAATTT CCTGTTCTAG CTGCTGCTTC AGTTGGCGGT CGACCACTTC TGCAACTAAG 4200

AGTTGACTGG CTGCAATGGG GGTCTTTATC TGATGGACCC ACAAGGTATA GTAATCCAGC 4260

AAATCCGTCA GTTTTCTTTC TGCTTTTGAC CTCTGCTGAT AGAGTTCCAT CTCACGCGCT 4320

TCTAATTTTT CTGCTAAAGC TATTTCCAAA GGAGACTTGG CTTCCCTCTC TCCATAGAGA 4380

AGTTCCTGGC GATAGACCTG CGTTTCCACC AATATGTCCC AAGTGAAAAA TAATATGGTT 4440

ACAAAGCAAC ACAAGAAGAA AAAGTAGAGG AAGTAAATTC CTAGACTGGC AAATAAAAAC 4500

TGAAAGAGTA AGACAAGAAA TGCCAAAGAA AGCAGATAGA TAAAAAGACG ACTACGGGAG 4560

CGCAGATAGG CTAGAAAAAA TTGTTTCCAA TCAAGCATGC TTCAATCCGT ACCCTATTCC 4620

TTTCTTGGTC TCGATAAATC CTACCAATCC CTGCTCCTCC AACTTTTTAC GCAAACGAGC 4680

CACATTGACA GAGAGGGTAT TATCATCAAT GAAAAAGTCA CTGTTCCAAA GTTCCCGCAT 4740

CAGGTCGTCA CGTGCTACGA TGTTGCCTGC ATGCTCAAAT AACACGCGTA AAATCTGGAA 4800

TTCATTCTTG GTCAAATTCA AGACTTGCCC TTGATAATGT AAATCCATGG ATTTGGTATT 4860

GAGGATAACA CCAGCATATT CCAGCAAACT CTCATCACGC CCAAACTCAT AGGAACGACG 4920

CAACAAGCCC TGAACCTTAG CTAAAAGAAC CTGCTGGTCA AAAGGCTTGG TCACAAAGTC 4980

ATCCGCCCCC ATATTGATTG CCATGACAAT ATCCATAGCC TGGTCTCTCG AAGAAAGAAA 5040

CATGATAGGT ACCTTGGAAA TCTTGCGGAT TTCCTGACAC CAGTGATAAC CATTAAACAA 5100

GGGCAAACCA ATATCCATGA GGACCAGATG AGGTTCCGAC TGAACAAATA GACTCAAAAC 5160

TTCCATAAAG TCTTCTACCA GGACCACTTC AAATCCCCAT TCAGAGAGCA TTTTCCCAAT 5220

CTGTTGACGA ATGACCTGAT CATCTTCTAT TAATAAAATC TTGTGCATGC GCTTCTCCTT 5280

TTCCATTATT ATAACAGATT TTTCCATGCT AGATGGTCTG AAACTGAATT TGAAATAGCC 5340

TGTTTTTAGC CAGTACAAAC AGGCTATGCT ACTAGCTAAT TTGAGGGAAA TTTGCTAAGA 5400

TAAATAAAAA GAAAGGAGCT CTTATGGCCA ATATTTTTGA CTATCTGAAA GATGTCGCAT 5460

ATGATTCTTA TTACGACCTT CCCTTGAATG AGTTAGACAT TCTAACCTTA ATAGAAATCA 5520

CCTACCTCTC CTTTGATAAT CTGGTCTCCA CACTTCCTCA ACGTCTTTTA GATCTAGCAC 5580

CTCAGGTTCC AAGAGATCCC ACCATGCTTA CTAGCAAAAA TCGCCTTCAA TTATTAGATG 5640

AATTGGCTCA ACACAAGCGC TTCAAAAATT GCAAACTCTC CCATTTTATC AACGACATCG 5700

ACCCTGAACT GCAAAAGCAA TTTGCGGCTA TGACTTATCG TGTCAGCCTC GATACCTATC 5760

TGATTGTCTT TCGTGGGACA GATGACAGTA TCATTGGCTG GAAGGAAGAT TTCCACCTGA 5820

CCTATATGAA GGAAATTCCT GCTCAAAAGC ACGCCCTTCG CTATTTAAAG AACTTTTTTG 5880

CCCATCATCC TAAGCAAAAG GTTATTCTAG CTGGGCATTC CAAGGGAGGA AATCTCGCTA 5940

TCTATGCTGC TAGCCAAATT GAGCAAAGTT TGCAAAATCA GATCACAGCA GTTTATACAT 6000

TTGATGCACC TGGTCTCCAT CAAGAATTGA CACAGACTGC GGGTTATCAA AGGATAATGG 6060

ATAGAAGCAA GATATTCATT CCACAAGGTT CCATTATCGG TATGATGCTG GAAATTCCTG 6120

CTCACCAAAT CATCGTTCAG AGTACTGCCC TGGGTGGCAT CGCCCAGCAC GATACCTTTA 6180

GTTGGCAGAT TGAGGACAAG CACTTCGTCC AACTGGATAA GACCAACAGT GATAGCCAGC 6240

AAGTAGACAC AACCTTTAAA GAATGGGTGG CCACAGTCCC TGACGAAGAA CTTCAGCTCT 6300

ACTTCGACCT CTTCTTTGGC ACTATTCTTG ATGCTGGTAT TAGCTCTATC AATGACTTGG 6360

CTTCCTTAAA GGCGCTTGAA TACATTCATC ATCTCTTTGT CCAAGCTCAA TCCCTCACTC 6420

CAGAAGAAAG AGAAACCTTG GGTCGCCTTA CCCAGTTATT GATTGATACT CGTTACCAGG 6480

CATGGAAAAA TAGATAATAC TCTTGAAAAT TAAATGTATA CAAAACAAAA GACCTAGAAT 6540

ACATACTTTC ATGTGCATTC TAAGTCTTTT TAAATAGAAT CTAATAGTCA ATAAAAATCA 6600

AAGAGCATTG AGAGATAATG GGGCTTGGAA CGTCCCTCTC GCTTCAACAA AATGACCCCA 6660

TTATAGATTA AAAAGATGCC ACTTAGAAAA AGCAAAAAAG GAAGTAAGAC AAAGGCAAAT 6720

ATATAAAAAG CTAACTGAAC ATTCTCGTAT CCATTTTTAT AAAAAAGGTA GGATAGATAA 6780

AAATAACTTG AAATGAGGGA TAATAAAAAT AATACTGGAT TCCACAAACT TCTATTATCC 6840

TTCCAAAATG ACACTATAAA GGCTAATACA ATTCCTATAA CGAGATACAT TTCTTACTCC 6900

TTTAATAGCT ACATTTTATC ATAATTATCC AAAGAAAAAA GAGGGCATTT ATCCCTCTTA 6960

ATCCTTCATC TGACTCTCTG CATCGGCCAC GACTTTTTCT AGACTGGTTT GACCAAGTTC 7020

TGCCTCCATA GTCAACTGAA TTCTCTCCAA TTTTTGATCC AAAACATCAT GAATATGAGC 7080

TCCTACAGGG CAATTTGGAT TCGGATTGTC ATGGAAACTG AAGAGTTGAC CTGTCTTACC 7140

AAGACATTCG ACCGCCTGAT AAACATCTAA AAGACTAATA TCCTTAAGGT CCTTGACAAT 7200

CTCTGTTCCG CCCGTTCCAC GCGCTACTGA AATCAGCTCT GCCTTCTTCA ACTGGGACAA 7260

GATCTTTCTG ATAATGACAG GATTGACCCC GACACTAGCA GCCAGAAAAT CACTGGTCAC 7320

CTTGCTTTCC TTCCCCTCGA GGGCAATGAT TATCAGCATA TGAGTCGCAA TGGTAAATCT 7380

ACTTGGAATT TGCATCCTCT TCTCCTTTTT ACGAGGCTAC CCTGCCTCTA CTCTTCTTTT 7440

TCTATTATTA TACCCTTTTT AGTTGTAATG TCAATCGTTA CCACTTTTCA ACCAGTCGTC 7500

TAACTCCCGA TCGCAGCCCT CTTTCTGAGC CAATTCTCTC AAAAATTCCT GATGATGAGT 7560

ATGGTGGATC CCATTGACCA GACTTTCATA GTAAACCTCA AAATAGGGAA GTCTCAGGTC 7620

TTTAGCCAGC TGCAATTCAG CTGCTACATC GTAGTCTACC CGTCGGAAGT CCATATCTAC 7680

CAGGCCTTTG TCATCAAACT CCAAAATCAT ATACTGGGCC CGCAAGTCCT TCCGTAGCTG 7740

AGCGTCCAAA AAGAAAGGTT GGCCAATCGA ACCCGGATTG ACAATCAATT GCCCACCAGT 7800

CCCGTAACGA AGCAACTGCT GGTGAATATC TCCATAAACA GCAATATCAC AGGGAGGATG 7860

AGTCACCAAG CGGTCAAACT CCTCTTGTTT GCCAGTATGA ATCAACTCTC GCCCCCAGTT 7920

CTTATCAGGC AGATGATGGC TAATTCCCAC CGTCAAATCC CCAAACTGAC GATGAATTTG 7980

AAGAGGTTGA TTGTGGAGCA CTTCAATTTC TTCTAGGGAA ATTTCCTCTA AAACATACTG 8040

GCACTGGCGC AAGAGATAGC GTTGACTGGG GCGAGTACTG TCCAATTCCT TACGGACACC 8100

ATGCCAAAGA CTGTCTTCCC AGTTTCCCAA AACTCTAGCC GTAATCGGTA GTTGATCCAA 8160

CAAGTCCAAA ATCCTTCTAC GCCCTGTCCC TGGCATGAGA ATATCTCCCA AAAGCCAGTA 8220

TTCATCCACT CCTATCTGCC GAGCATCTGC CAAAACAGCC TCCAAGGCGG TGGTATTTCC 8280

ATGAATATCT GAAAGAAGAG CTATTTTCGT CATATCCATC TCCTCGTTTT TTCTCTTGCA 8340

ATAAGTATAA CATAAAAAGT CACAGCTAGA GAAATCTAGC TTTTTTTGAT ATACTAGATA 8400

AAGATATTAG ACAAGAGGAA ACGAATGACC CCAAACAAAG AAGACTATCT AAAATGTATT 8460

TATGAAATTG GCATAGACCT GCATAAGATT ACCAACAAGG AAATTGCGGC TCGCATGCAA 8520

GTCTCTCCCC CTGCCGTAAC TGAAATGATC AAACGAATGA AAAGTGAAAA TCTCATCCTA 8580

AAGGACAAGG AATGTGGCTA TCTACTGACT GACCTCGGTC TCAAACTGGT CTCTGAGCTC 8640

TATCGTAAGC ACCGCTTGAT TGAAGTTTTT CTAGTTCATC ATTTAGACTA TACAAGTGAC 8700

CAGATTCACG AGGAAGCTGA GGTCTTGGAA CACACTGTCT CTGACCTGTT CGTGGAAAGA 8760

CTAGATAAAC TGCTAGGTTT CCCTAAAACC TGCCCCCACG GGGGAACTAT TCCTGCCAAG 8820

GGAGAACTAC TCGTTGAAAT CAATAACCTC CCACTAGCTG ATATCAAGGA AGCTGGCGCC 8880

TACCGCCTGA CTCGGGTGCA CGATAGTTTT GACATTCTCC ATTATCTGGA CAAGCACTCA 8940

CTTCACATCG GTGACCAGCT CCAAGTCAAG CAGTTTGATG GCTTCAGCAA TACCTTCACT 9000

ATCCTCAGTA ACGACGAGGA TTTACAAGTG AATATGGACA TTGCAAAACA ACTCTATGTC 9060

GAGAAAATCA ACTAATTTCT CAAGTCCCCT ACCAACCCTG AAAGTTTTAT TTTGGCTCTT 9120

TGTCAACTGT AGTGGGTTGA AGTCAGCTAA GCTCGAGAAA GGACAAATTT TGTCCTTTCT 9180

TTTTTGATAT TCAGAGCGAT AAAAATCCGT TTTTTGAAGT TTTCAAAGTT CCGAAAACCA 9240

AAGGCATTGC GCTTGATAAG TTTGATGAGA TTATTGGTCG CTTCCAGTTT GGCATTAGAA 9300

TAGTGTAGTT GAAGGGCGTT GACAATCTTT TCTTTATCTT TGAGGAAGGT TTTAAAGACA 9360

GTCTGAAAAA TAGGATGAAC CTGCTTTAGA TTGTCCTCAA TGAGTCCGAA AAATTTCTCC 9420

GGTTTCTTAT TCTGAAAGTG AAACAGCAAG AGTTGATAGA GCTGATAGTG GTGTTTCAAG 9480

TCTTGTGAAT AGCTCAAAAG CTTGTCTAAA ATCTCTTTAT TGGTTAAGTG CATACGAAAA 9540

GTAGGACGAT AAAATCGCTT ATCACTCAGT TTACGGCTAT CCTGTTGTAT GAGCTTCCAG 9600

TAGCGCTTGA TAGCCTTGTA TTCATGGGAT TTTCGATCCA ATTGGTTCAT AATTTGAACA 9660

CGCACACGAC TCATAGCACG GCTAAGATGT TGTACAATGT GAAAGCGATC CAACACGATT 9720

TTAGCATTCG GGAGTGAAAC AGTCTGGGAG ACTGTTTCAG CCTGAGCCTA GAAATTTGAA 9780

AGCGAAGCTG TTTAGCCAAG TCATAGTAAG GACTAAACAT ATCCATCGTA ATGATTTTCA 9840

CTTGACAACG AACGGCTCTA TCGTAGCGAA GAAAGTGATT TCGGATGACA GCTTGTGTTC 9900

TGCCTTCAAG AACAGTGATA ATATTAAGAT TATCAAAATC TTGCGCAATG AAACTCATCT 9960

TTCCCTTAGT GAAGGCATAC TCATCCCAAG ACATAATCTT TGGAAGCCGA GAAAAATCAT 10020

GCTCAAAGTG AAAGTCATTG AGCTTGCGAA TGACAGTTGA AGTTGAAATG GCCAGCTGAT 10080

GGGCAATATC AGTCATAGAA ATTTTTTCAA TTAACTTTTG AGCAATyTTT TGGTTGATGA 10140

TACGAGGGAT TTGGTGATTT TTCTTTACCA GGGGAGTCTC AGCAACCATC ATTTTTGAAC 10200

AGTGATAGCA CTTGAAACGA CGCTTTCTAA GGAGAATTCT AGAAGGCATA CCAGTCGTTT 10260

CAAGATAAGG AATTTTAGAA GGTTTTTGAA AGTCATATTT CTTCAATTGG TTTCCGCACT 10320

CAGGGCAAGA TGGGGCGTCG TAGTCCAGTT TGGCGATGAT TTCCTTGTGT GTATCCTTAT 10380

TGATGATGTC TAAAATCTGG ATATTAGGGT CTTTAATGTC TAGTAATTTT GTGATAAAAT 10440

GTAATTGTTC CATATGATTC TTTCTAATGA GTTGTTTTGT CGCTTTTCAT TATAGGTCAT 10500

ATGGGACTTT TTTTCTACAA TAAAATAGGC TCCATAATAT CTATAGTGGA TTTACCCACT 10560

ACAAATATTA TAGAACCGTA AAAATAGAAG GAGATAGCAG GTTTTCAAGC CTGCTATCTT 10620

TTTTTGATGA CATTCAGGCT GATACGAAAT CATAAGAGGT CTGAAACTAC TTTCAGAGTA 10680

GTCTGTTCTA TAAAATATAG TAGATTGAAA TAAGATGTGA ACAACTCTAT CAGGAAAGTC 10740

AAATTAATTT ATAGAATTAT TTTAGCAGTC AAGGTGTACT GTTATAGATT CAATATATTA 10800

TATGACTATT AACCTTGTCT TCTCCTAAAA TTGACTTTCT TGTTTTCTTA TCTTGTCCAC 10860

TCGAAACAAG TATTGTAAGA ATTTGATTAT TTTTGAAAGT ACTTTTAATA TACTTGATAT 10920

AGTTAAAAAA GATTTGAAAC TAAATTCCAA ATTAGAAAAA GACTTGAAAT ACTAAAAAAA 10980

AAAAAGTATA CTCTAATTGA AAACGGTAAC AAAACTAATT TAGAGAATGA AATATAGAGT 11040

ATTTCTCTCT TAAAAGTTTT TGGTGAAACG AGATGTAGAA AGGAGATTTA GCCAAAGAGT 11100

CTATTAGTGC TAGAATAATA GATTAGAATT ATTTTAGAAA AACGAAGTGA GCAGCTTATA 11160

AATTCAAGTC CCCAAATAGA TTCATACTAG TATCTTTTGC AAAAAATAAA GGGCGACTTC 11220

CTTCATGAAT ATCAATTTCA TCTATAAGGA AGGTAGCTAA TTGAACTAAC TTATTTATTC 11280

TGTTTGTCGC TAGAAAAATC AGACCTCCTT GTGAAGATTG AGGAGATACT TAATGAAAAT 11340

CAAAGAAGAA ACTAGCAAGC TAGTAGCAGA TTGCCCAAAA CACCGCTTTG AGGTTGTAGA 11400

TAAGACTGAC CTATATAATC CAAGGTGAAG CGACTGTGGT TTGAAGAGAT TTTCAAAGAG 11460

TATAGGCTAG AGAGTAGTGT TTTTATGTCC TTCTAGTAGA AAATGCTAGA CAGAAGAATG 11520

GGGAACTTGG ATAGGAAAAA TAGATTGAGA AAGGAGGTTA GAAGAGATGA TTATTACAAA 11580

AATTAGCCGT TTAGGAACTT ATGTGGGAGT AAATCCACAT TTTGCAACAT TAATAGATTT 11640

TCTAGAAAAA ACAGGACTAG AAAATTTAAC AGAAGGTTCG ATTGCTATCG ATGGTAATCG 11700

ATTGTTTGGG AATTGCTTTA CTTATCTAGC AGATGGTCAA GCAGGGGCTT TCTTTGAAAC 11760

CCACCAAAAA TATTTGGATA TTCATTTAGT TTTGGAAAAC GAAGAAGCCA TGGCTGTTAC 11820

ATCGCCGGAA AATGTAAGCG TTACCCAAGA ATATGATGAA GAGAAAGATA TTGAATTATA 11880

CACAGGGAAA GTGGAACAGT TGGTTCATTT GAGAGCTGGC GAATGCCTCA TCACTTTTCC 11940

AGAAGATTTA CATCAACCCA AGGTTCGTAT AAATGATGAA CCTGTGAAAA AAGTTGTCTT 12000

TAAAGTTGCG ATTTCTTAAT GTAGAAAGAG AAGAACGATG AAAAAAATGA GAAAGTTTTT 12060

ATGTCTAGCT GGAATTGCGC TAGCGGCTGT TGCCTTGGTA GCTTGTTCAG GAAAAAAAGA 12120

AGCTACAACT AGTACTGAAC CACCAACAGA ATTATCTGGT GAGATTACAA TGTGGCACTC 12180

CTTTACTCAA GGACCCCGTT TAGAAAGTAT TCAAAAATCA GCAGATGCTT TCATGCAAAA 12240

GCATCCAAAA ACGAAAATCA AGATTGAAAC ATTTTCTTGG AATGACTTCT ATACTAAATG 12300

GACTACAGGT TTAGCAAATG GAAATGTGCC AGATATCAGT ACAGCTCTTC CTAACCAAGT 12360

AATGGAAATG GTCAACTCAG ATGCTTTGGT TCCGCTAAAT GATTCTATCA AGCGTATTGG 12420

ACAAGATAAA TTTAACGAAA CTGCCTTAAA TGAAGCAAAA ATCGGAGATG ATTACTACTC 12480

TGTTCCTCTT TATTCACATG CACAAGTCAT GTGGGTTAGA ACAGATTTGT TAAAAGAACA 12540

TAATATTGAG GTTCCTAAAA CTTGGGATCA ACTCTATGAA GCTTCTAAAA AATTGAAAGA 12600

AGCTGGAGTT TATGGCTTGT CTGTTCCGTT TGGAACAAAT GACTTAATGG CAACACGTTT 12660

CTTGAACTTC TACGTACGTA rTGGTGGAGG AAGCCTCTTA ACAAAAGATC TTAAAGCAGA 12720

CTTGACAAGC CAACTTGCTC AAGATGGTAT TAAATACTGG GTTAAATTGT ATAAAGAAAT 12780

CTCACCTCAA GATTCTTTGA ACTTTAATGT CCTTCAACAA GCTACCTTGT TCTATCAAGG 12840

AAAAACAGCA TTTGACTTTA ACTCTGGCTT CCATATCGGA GGAATTAATG CCAACAGTCC 12900

TCAATTGATT GATTCGATTG ATGCTTATCC TATTCCAAAA ATCAAAGAGT CTGATAAAGA 12960

CCAAGGAATT GAAACCTCAA ACATTCCAAT GGTTGTTTGG AAAAATTCAA AACATCCAGA 13020

AGTTGCTAAA GCATTCTTAG AAGCACTTTA TAATGAAGAA GACTACGTTA AATTCCTTGA 13080

TTCAACTCCA GTAGGTATGT TGCCAACTAT TAAGGGGATT AGCGATTCTG CAGCCTATAA 13140

AGAAAATGAA ACTCGTAAGA AATTTAAACA TGCTGAAGAA GTAATTACTG AAGCTGTTAA 13200

AAAAGGTACT GCTATTGGTT ATGAAAATGG GCCAAGTGTA CAAGCTGGTA TGTTGACTAA 13260

CCAACACATT ATTGAACAAA TGTTCCAAGA TATCATTACA AATGGAACAG ATCCTATGAA 13320

AGCAGCAAAA GAAGCAGAAA AACAATTAAA TGATTTATTT GAGGCTGTTC AGTAGATGTA 13380

AAAGACTAGA AAATAGGTGG GATAGTGAGC TGAAAAGCTC TAGCCCAATC TTGTAAAAGA 13440

AGGGAGAAGG AGAATGGTTA AAGAACGTAA TTTAACTCGC TGGATATTTG TTTTGCCAGC 13500

TATGATTATC GTAGGATTAC TCTTTGTTTA TCCGTTTTTC TCGAGTATTT TTTATAGCTT 13560

TACCAATAAG CATTTGATTA TGCCTAATTA TAAATTTGTT GGTTTGGCTA ACTATAAAGC 13620

TGTGCTATCA GATCCCAACT TCTTTAATGC GTTCTTTAAT TCAATTAAGT GGACCGTTTT 13680

CTCATTAGTT GGTCAAGTTT TAGTAGGGTT TGTATTGGCT TTAGCTCTTC ACAGAGTACG 13740

CCACTTCAAG AAATTATATA GGACATTATT GATTGTTCCT TGGGCATTTC CTACCATCGT 13800

TATTGCCTTC TCTTGGCAGT GGATTCTAAA CGGGGTTTAT GGCTACTTAC CTAATCTAAT 13860

CGTAAAATTA GGTTTAATGG AACATACACC TGCATTTTTG ACAGATAGTA CATGGGCATT 13920

CCTATGTTTG GTGTTTATCA ACATTTGGTT TGGAGCACCA ATGATTATGG TTAATGTGCT 13980

TTCAGCTTTG CAAACAGTAC CAGAAGAACA ATTTGAGGCT GCTAAGATAG ATGGTGCTTC 14040

AAGTTGGCAG GTGTTCAAGT TTATCGTCTT TCCACATATT AAAGTGGTTG TAGGACTTCT 14100

AGTTGTTTTG AGAACTGTAT GGATCTTTAA TAACTTTGAC ATTATCTACC TCATTACTGG 14160

TGGTGGACCA GCCAATGCTA CAACGACGCT TCCAATTTTT GCTTACAACC TGGGCTGGGG 14220

AACTAAATTG TTGGGTCGTG CTTCAGCAGT TACAGTACTG CTCTTTATCT TCTTGGTGGC 14280

GATTTGCTTT ATCTACTTTG CTATCATCAG TAAGTGGGAA AAGGAGGGTA GAAAATAATG 14340

AAGAAGAAAT CCAGTATTTA TTTAGATATT CTCTCACATG TACTTTTAGT TGGTGCGACC 14400

ATCGTTGCAG TTTTCCCATT GGTATGGATT ATCATATCTT CTGTCAAAGG GAAAGGGGAA 14460

TTAACTCAGT ATCCAACACG ATTTTGGCCT GAACAGTTTA CATTAGATTA TTTCACTCAT 14520

GTTATCAACG ATTTGCACTT CATTGATAAC ATTCGAAACA GTTTAATCAT TGCCTTGGCT 14580

ACAACCCTTA TTGCGATTAT TATTTCTGCT ATGGCAGCCT ATGGTATTGT TCGATTCTTT 14640

CCTAAATTGG GAGCAATCAT GTCGAGACTA CTCGTCATTA CCTACATTTT CCCACCAATT 14700

TTGTTAGCAA TTCCCTATTC AATTGCCATT GCTAAAGTTG GGTTAACAAA TAGTTTATTT 14760

GGCTTGATGA TGGTTTATCT ATCTTTTAGT GTTCCATATG CAGTTTGGCT CTTAGTTGGA 14820

TTTTTCCAAA CAGTTCCAAT TGGAATTGAA GAAGCGGCTA GAATTGATGG TGCAAATAAA 14880

TTTGTTACGT TTTATAAAGT TGTGCTACCG ATTGTAGCAC CAGGTATTGT AGCAACAGCT 14940

ATTTATACAT TTATCAATGC TTGGAATGAA TTCCTGTATG CCTTGATTTT GATTAACAAT 15000

ACAGGAAAGA TGACAGTAGC AGTAGCCCTT CGTTCACTTA ATGGTTCAGA AATACTAGAC 15060

TGGGGAGATA TGATGGCAGC GTCTGTTATT GTAGTTCTTC CATCAATTAT TTTCTTCTCT 15120

ATCATCCAAA ATAAGATTGC AAGTGGATTA TCAGAAGGAT CTGTGAAGTA GACGAAAGAA 15180

GGAAAAAAAT GAATAAAAGA GGTCTTTATT CAAAACTAGG AATTTCCGTT GTAGGCATTA 15240

GTCTTTTAAT GGGAGTCCCC ACTTTGATTC ATGCGAATGA ATTAAACTAT GGTCAACTGT 15300

CCATATCTCC TATTTTTCAA GGAGGTTCAT ATCAACTGAA CAATAAGAGT ATAGATATCA 15360

GCTCTTTGTT ATTAGATAAA TTGTCTGGAG AGAGTCAGAC AGTAGTAATG AAATTTAAAG 15420

CAGATAAACC AAACTCTCTT CAAGCTTTGT TTGGCCTATC TAATAGTAAA GCAGGCTTTA 15480

AAAATAATTA CTTTTCAATT TTCATGAGAG ATTCTGGTGA GATAGGTGTA GAAATAAGAG 15540

ACGCCCAAAA GGGAATAAAT TATTTATTTT CCAGACCAGC TTCATTATGG GGAAAACATA 15600

AAGGACAGGC AGTTGAAAAT ACACTAGTAT TTGTATCTGA TTCTAAAGAT AAAACATACA 15660

CAATGTATGT TAATGGAATA GAAGTGTTCT CTGAAACAGT TGATACATTT TTGCCAATTT 15720

CAAATATAAA TGGTATAGAT AAGGCAACAC TAGGAGCTGT TAATCGTGAA GGTAAGGAAC 15780

ATTACCTCGC AAAAGGAAGT ATTGATGAAA TCAGTCTATT TAACAAAGCA ATTAGTGATC 15840

AGGAAGTTTC AACTATTCCC TTGTCAAATC CATTTCAGTT AATTTTCCAA TCAGGAGATT 15900

CTACTCAAGC TAACTATTTT AGAATACCGA CACTATATAC ATTAAGTAGT GGAAGAGTTC 15960

TATCAAGTAT TGATGCACGT TATGGTGGGA CTCATGATTC TAAAAGTAAG ATTAATATTG 16020

CCACTTCTTA TAGTGATGAT AATGGGAAAA CGTGGAGTGA GCCAATTTTT GCTATGAAGT 16080

TTAATGACTA TGAGGAGCAG TTAGTTTACT GGCCACGAGA TAATAAATTA AAGAATAGTC 16140

AAATTAGTGG AAGTGCTTCA TTCATAGATT CATCCATTGT TGAAGATAAA AAATCTGGGA 16200

AAACGATATT ACTAGCTGAT GTTATGCCTG CGGGTATTGG AAATAATAAT GCAAATAAAG 16260

CCGACTCAGG TTTTAAAGAA ATAAATGGTC ATTATTATTT AAAACTAAAG AAGAATGGAG 16320

ATAACGATTT CCGTTATACA GTTAGAGAAA ATGGTGTCGT TTATAATGAA ACAACTAATA 16380

AACCTACAAA TTATACTATA AATGATAAGT ATGAAGTTTT GGAGGGAGGA AAGTCTTTAA 16440

CAGTCGAACA ATATTCGGTT GATTTTGATA GTGGCTCTTT AAGAGAAAGG CATAATGGAA 16500

AACAGGTTCC TATGAATGTT TTCTACAAAG ATTCGTTATT TAAAGTGACT CCTACTAATT 16560

ATATAGCAAT GACAACTAGT CAGAATAGAG GAGAGAGTTG GGAACAATTT AAGTTGTTGC 16620

CTCCGTTCTT AGGAGAAAAA CATAATGGAA CTTACTTATG TCCCCGGACAA GGTTTAGCAT 16680

TAAAATCAAG TAACAGATTG ATTTTTGCAA CATATACTAG TGGAGAACTA ACCTATCTCA 16740

TTTCTGATGA TAGTGGTCAA ACATGGAAGA AATCCTCAGC TTCAATTCCG TTTAAAAATG 16800

CAACAGCAGA AGCACAAATG GTTGAACTGA GAGATGGTGT GATTAGAACA TTCTTTAGAA 16860

CCACTACAGG TAAGATAGCT TATATGACTA GTAGAGATTC TGGAGAAACA TGGTCGAAAG 16920

TTTCGTATAT TGATGGAATC CAACAAACTT CATATGGCAC ACAAGTATCT GCAATTAAAT 16980

ACTCTCAATT AATTGATGGA AAAGAAGCAG TCATTTTGAG TACACCAAAT TCTAGAAGTG 17040

GCCGCAAGGG AGGCCAATTA GTTGTCGGTT TAGTCAATAA AGAAGATGAT AGTATTGATT 17100

GGAAATACCA CTATGATATT GATTTGCCTT CGTATGGTTA TGCCTATTCT GCGATTACAG 17160

AATTGCCAAA TCATCACATA GGTGTACTGT TTGAAAAATA TGATTCGTGG TCGAGAAATG 17220

AATTGCATTT AAGCAATGTA GTTCAGTATA TAGATTTGGA AATTAATGAT TTAACAAAAT 17280

AAAGGAGAAA AACATGGTTA AATACGGTGT TGTTGGAACA GGGTATTTTG GAGCTGAATT 17340

GGCTCGCTAC ATGCAAAAGA ATGATGGAGC AGAGATTACT CTTCTCTATG ATCCAGATAA 17400

TGCAGAGGCG ATTGCAGAAG AATTGGGAGC AAAAGTAGCA AGTTCCTTAG ATGAGTTGGT 17460

TTCTAGCGAT GAAGTAGATT GTGTTATCGT CGCAACTCCA AATAATCTTC ATAAGGAACC 17520

GGTTATTAAG GCTGCACAGC ATGGTAAAAA TGTTTTCTGT GAAAAACCAA TTGCGCTTTC 17580

TTATCAAGAT TGTCGCGAGA TGGTAGATGC GTGTAAAGAA AACAATGTAA CCTTTATGGC 17640

AGGACATATT ATGAATTTCT TTAATGGTGT TCATCATGCA AAAGAACTCA TTAATCAAGG 17700

AGTTATCGGA GACGTTCTAT ATTGTCATAC AGCTCGTAAT GGTTGGGAAG AACAACAACC 17760

GTCAGTATCA TGGAAAAAAA TTCGTGAAAA ATCAGGTGGT CACTTGTATC ACCACATCCA 17820

TGAATTGGAT TGCGTTCAAT TCCTTATGGG GGGCATGCCT GAAACTGTAA CCATGACAGG 17880

TGGAAATGTG GCCCATGAAG GTGAACATTT CGGTGATGAA GATGATATGA TTTTTGTCAA 17940

TATGGAATTT TCTAATAAGC GTTTTGCCTT GTTAGAATGG GGTTCAGCTT ATCGTTGGGG 18000

TGAACATTAT GTCTTAATCC AAGGAAGCAA AGGTGCCATC CGCTTAGACT TATTCAACTG 18060

TAAAGGAACT CTTAAGCTAG ATGGGCAAGA AAGCTATTTC TTGATTCACG AATCGCAAGA 18120

AGAAGATGAT GATCGGACTC GTATCTATCA TAGTACAGAG ATGGATGGAG CAATTGCTTA 18180

TGGTAAACCA GGTAAACGTA CTCCATTATG GCTATCATCT GTCATTGATA AAGAAATGCG 18240

CTATCTGCAT GAGATTATGG AAGGAGCTCC AGTATCAGAA GAATTTGCAA AACTTTTGAC 18300

AGGTGAAGCT GCCCTAGAAG CAATTGCTAC TGCAGATGCT TGTACCCAGT CTATGTTTGA 18360

AGATCGCAAA GTAAAATTGT CAGAAATTGT AAAATAAATT TTGGTATTCT CCTATTTATA 18420

GGTCGACTTG CTCCTCTGAA AGTACTTTTA GAGGAGCTGT TTGACTTTGC TAGTTTTTGA 18480

AACTGAAATC TATTATACTA CAAACTATTG AAAGCGTTTT AATTTTAAGG TATAATAATC 18540

TCATAGAAAT AAAGAAAAGG AGGAAAGAGG ATGCCACAGA TTAGCAAAGA AGCCTTGATT 18600

GAGCAAATCA AAGATGGAAT CATCGTTTCT TGTCAGGCTC TTCCTCATGA ACCGCTTTAT 18660

ACAGAAGCGG GAGGGGTGAT TCCCTTGCTG GTCAAAGCGG CTGAGCAAGG TGGAGCAGTC 18720

GGTATCCGAG CAAACAGTGT TCGCGATATC AAGGAAATTA AGGAAGTCAC TAAACTTCCA 18780

ATCATTGGGA TTATCAAACG TGATTATCCA CCTCAGGAAC CCTTCATCAC GGCTACTATG 18840

AAAGAAGTTG ATGAATTGGC AGAACTGGAC ATCGAGGTGA TTGCTCTGGA TTGTACCAAG 18900

CGTGAACGCT ACGATGGTTT GGAAATTCAA GAGTTCATTC GTCAGGTTAA GGAGAAATAT 18960

CCTAATCAGC TTTTGATGGC TGATACTAGT ATCTTCGAAG AAGGGCTAGC AGCTGTAGAA 19020

GCAGGAATTG ACTTTGTCGG AACAACCTTA TCAGGCTACA CATCCTACAG TCCAAAAGTA 19080

GACGGTCCAG ATTTTGAATT GATTAAGAAA CTCTGTGATG CTGGTGTAGA TGTCATTGCA 19140

GAAGGAAAAA TTCATACACC AGAACAAGCC AAACAAATCC TTGAATATGG AGTGCGAGGC 19200

ATCGTTGTTG GTGGCGCCAT TACTAGACCA AAAGAGATTA CAGAACGCTT CGTTGCTAGT 19260

CTTAAATAAG ATGTGAGGGG GAGTTTTATG TTTAAAGTTT TACAAAAAGT TGGAAAAGCT 19320

TTTATGTTAC CTATAGCTAT ACTTCCTGCA GCAGGTCTAC TTTTGGGGAT TGGTGGTGCA 19380

CTTTCAAACC CAACCACGAT AGCAACTTAT CCAATACTAG ACAATAGTAT TTTTCAATCA 19440

ATATTCCAAG TAATGAGCTC TGCAGGAGAG GTTGTATTCA GTAATTTGTC ACTACTTCTC 19500

TGTGTGGGAT TATGTATTGG CTTAGCGAAA CGAGATAAAG GAACCGCTGC GTTAGCAGGA 19560

GTAACTGGTT ACTTAGTTAT GACTGCAACG ATCAAAGCTT TGGTAAAACT TTTTATGGCA 19620

GAAGGATCTG CAATTGATAC TGGAGTTATT GGAGCATTAG TTGTCGGAAT AGTTGCCGTA 19680

TATTTGCACA ACCGATATAA CAATATTCAA TTACCTTCCG CTTTAGGATT CTTTGGAGGT 19740

TCACGCTTCG TTCCTATTGT TACATCGTTC TCTTCTATCT TGATTGGCTT TGTCTTCTTT 19800

GTTATTTGGC CACCTTTCCA ACAACTTCTT GTTTCTACAG GTGGATATAT TTCTCAGGCG 19860

GGTCCAATTG GAACTTTTCT ATATGGATTT TTAATGAGAC TTTCTGGAGC AGTAGGCTTA 19920

CATCATATAA TTTACCCTAT GTTTTGGTAT ACTGAACTTG GTGGTGTTGA AACTGTTGCA 19980

GGACAAACAG TGGTTGGAGC TCAAAAAATA TTTTTTGCTC AATTAGCCGA TTTGGCCCAT 20040

TCTGGATTAT TTACAGAAGG AACAAGGTTT TTTGCAGGTC GTTTCTCAAC AATGATGTTC 20100

GGTTTACCGG CTGCCTGTTT AGCGATGTAC CATAGTGTTC CTAAAAATCG TCGTAAAAAA 20160

TACGCGGGTT TGTTTTTTGG AGTTGCTTTA ACATCTTTTA TTACCGGTAT TACAGAACCA 20220

ATTGAATTTA TGTTTCTATT CGTCAGTCCG GTTCTATATG TTGTTCACGC ATTCCTTGAT 20280

GGTGTTAGCT TCTTTATTGC AGACGTCTTA AATATTTCAA TAGGAAACAC ATTTTCAGGA 20340

GGTGTAATCG ATTTCACTTT ATTTGGAATT TTGCAGGGGA ACGCTAAGAC GAATTGGGTT 20400

CTTCAGATTC CATTTGGACT TATTTGGAGT GTTTTGTATT ATATTATTTT TAGATGGTTC 20460

ATTACTCAAT TCAACGTTCT AACGCCAGGG CGAGGAGAAG AAGTAGATTC TAAAGAAATT 20520

TCTGAATCCG CAGATTCAAC TTCAAATACT GCAGATTATT TAAAACAGGA TAGCCTACAA 20580

ATTATCAGAG CCTTGGGTGG ATCAAATAAT ATAGAAGATG TAGATGCTTG TGTGACACGT 20640

TTACGTGTAG CTGTAAAAGA AGTTAATCAA GTTGATAAAG CACTTTTAAA ACAAATTGGT 20700

GCAGTTGATG TCTTAGAAGT GAAGGGTGGC ATTCAAGCAA TCTATGGAGC AAAAGCAATC 20760

TTATATAAAA ATAGTATTAA TGAAATTTTA GGTGTAGATG ATTAAGTACT TACTGACTTA 20820

ATAAAAAAAC GAGGAGAGTG ATGGATGAGT AGGATGAAAT GAAATCGCAT ACAAGAAATA 20880

AAGAACTCAT TATCCAAGTT GGATACGCTT ATTACATAGG AGAATACAAA TGAAATTTAG 20940

AAAATTAGCT TGTACAGTAC TTGCGGGTGC TGCGGTTCTT GGTCTTGCTG CTTGTGGCAA 21000

TTCTGGCGGA AGTAAAGATG CTGCCAAATC AGGTGGTGAC GGTGCCAAAA CAGAAATCAC 21060

TTGGTGGGCA TTCCCAGTAT TTACCCAAGA AAAAACTGGT GACGGTGTTG GAACTTATGA 21120

AAAATCAATC ATCGAAGCGT TTGAAAAAGC AAACCCAGAT ATAAAAGTGA AATTGGAAAC 21180

CATCGACTTC AAGTCAGGTC CTGAAAAAAT CACAACAGCC ATCGAAGCAG GAACAGCTCC 21240

AGACGTACTC TTTGATGCAC CAGGACGTAT CATCCAATAC GGTAAAAACG GTAAATTGGC 21300

TGAGTTGAAT GACCTCTTCA CAGATGAATT TGTTAAAGAT GTCAACAATG AAAACATCGT 21360

ACAAGCAAGT AAAGCTGGAG ACAAGGCTTA TATGTATCCG ATTAGTTCTG CCCCATTCTA 21420

CATGGCAATG AACAAGAAAA TGTTAGAAGA TGCTGGAGTA GCAAACCTTG TAAAAGAAGG 21480

TTGGACAACT GATGATTTTG AAAAAGTATT GAAAGCACTT AAAGACAAGG GTTACACACC 21540

AGGTTCATTG TTCAGTTCTG GTCAAGGGGG AGACCAAGGA ACACGTGCCT TTATCTCTAA 21600

CCTTTATAGC GGTTCTGTAA CAGATGAAAA AGTTAGCAAA TATACAACTG ATGATCCTAA 21660

ATTCGTCAAA GGTCTTGAAA AAGCAACTAG CTGGATTAAA GACAATTTGA TCAATAATGG 21720

TTCACAATTT GACGGTGGGG CAGATATCCA AAACTTTGCC AACGGTCAAA CATCTTACAC 21780

AATCCTTTGG GCACCAGCTC AAAATGGTAT CCAAGCTAAA CTTTTAGAAG CAAGTAAGGT 21840

AGAAGTGGTA GAAGTACCAT TCCCATCAGA CGAAGGTAAG CCAGCTCTTG AGTACCTTGT 21900

AAACGGGTTT GCAGTATTCA ACAATAAAGA CGACAAGAAA GTCGCTGCAT CTAAGAAATT 21960

CATCCAGTTT ATCGCAGATG ACAAGGAGTG GGGACCTAAA GACGTAGTTC GTACAGGTGC 22020

TTTCCCAGTC CGTACTTCAT TTGGAAAACT TTATGAAGAC AAACGCATGG AAACAATCAG 22080

CGGCTGGACT CAATACTACT CACCATACTA CAACACTATT GATGGATTTG CTGAAATGAG 22140

AACACTTTGG TTCCCAATGT TGCAATCTGT ATCAAATGGT GACGAAAAAC CAGCAGATGC 22200

TTTGAAAGCC TTCACTGAAA AAGCGAACGA AACAATCAAA AAAGCTATGA AACAATAGTC 22260

CTTAGTTATT CTATAAAAAG TAGTTTTTTA AAGAACCTAA GAGTGTATAC CCCCTTTTCC 22320

CTCTACACAG ATAGTGTAAG AAAAGGGGGC TTTTGTTTAA AATGTAAGAA ACTGTCACGA 22380

AATTAAAATG AAGTTCTTAC ATAAGCGAAT CATAAAAAAT TTCATTTTGA TTTTAAAACA 22440

GTTCAAGAAA GTCAAAAAAT TATTCTATTT GAAAGAGAGG TGCCGACTGT GAAAGTCAAT 22500

AAAATCCGTA TGCGGGAAAC AGTGATTTCC TACGCTTTCC TAGCACCAGT ATTATTCTTC 22560

TTTGTCATCT TTGTGTTGGC TCCGATGGTG ATGGGCTTCA TTACAAGTTT CTTTAACTAC 22620

TCAATGACTA AATTTGAGTT TGTAGGCTTG GATAACTATA TCCGTATGTT TAAAGATCCT 22680

GTCTTTACAA AATCTCTGAT TAACACAGTT ATTTTGGTTA TTGGATCTGT ACCAGTTGTT 22740

GTTCTATTCT CACTCTTTGT AGCATCTCAG ACCTATCATC AAAATGTCAT TGCCAGATCC 22800

TTCTACCGTT TCGTCTTCTT CCTTCCTGTT GTAACGGGTA GTGTTGCCGT GACAGTTGTT 22860

TGGAAATGGA TTTATGACCC ACTATCAGGG ATTCTAAACT TTGTCCTTAA GTCCAGCCAC 22920

ATCATCAGCC AAAACATTTC TTGGTTGGGA GATAAAAACT GGGCATTGAT GGCGATTATG 22980

ATTATTCTCT TGACCACTTC AGTTGGTCAG CCCATCATCC TTTATATCGC TGCCATGGGG 23040

AATATTGACA ATTCACTGGT TGAAGCGGCG CGTGTTGATG GTGCAACTGA GTTTCAAGTT 23100

TTTTGGAAGA TTAAATGGCC AAGCCTTCTT CCAACAACTC TTTATATTGC AATCATCACA 23160

ACAATTAACT CATTCCAGTG TTTCGCCTTG ATTCAGCTTT TGACATCTGG TGGTCCAAAC 23220

TACTCAACAA GTACCTTGAT GTACTACCTT TACGAAAAAG CCTTCCAATT GACAGAATAC 23280

GGCTATGCCA ACACAATTGG TGTCTTCTTG GCAGTCATGA TTGCTATCGT AAGCTTTGTT 23340

CAATTTAAAG TACTTGGAAA CGACGTAGAA TACTAAAGAA AGGAGACAGC TATGCAATCT 23400

ACAGAAAAAA AACCATTAAC AGCCTTTACT GTTATTTCAA CAATCATTTT GCTCTTGTTG 23460

ACTGTGCTGT TCATCTTTCC ATTCTACTGG ATTTTGACAG GGGCATTCAA ATCACAACCT 23520

GATACAATTG TTATTCCTCC TCAGTGGTTC CCTAAAATGC CAACCATGGA AAACTTCCAA 23580

CAACTCATGG TGCAGAACCC TGCCTTGCAA TGGATGTGGA ACTCAGTATT TATCTCATTG 23640

GTAACCATGT TCTTAGTTTG TGCAACCTCA TCTCTAGCAG GTTATGTATT GGCTAAAAAA 23700

CGTTTCTATG GTCAACGCAT TCTATTTGCT ATCTTTATCG CTGCTATGGC GCTTCCAAAA 23760

CAAGTTGTCC TTGTACCATT GGTACGTATC GTCAACTTCA TGGGAATCCA TGATACTCTC 23820

TGGGCAGTTA TCTTGCCTTT GATTGGATGG CCATTCGGTG TCTTCCTCAT GAAACAGTTC 23880

AGTGAAAATA TCCCTACAGA GTTGCTTGAA TCAGCTAAAA TCGACGGTTG TGGTGAGATT 23940

CGTACCTTCT GGAGTGTAGC CTTCCCGATT GTGAAACCAG GGTTTGCAGC CCTTGCAATC 24000

TTTACCTTCA TCAATACTTG GAATGACTAC TTCATGCAAT TGGTAATGTT GACTTCACGT 24060

AACAATTTGA CCATCTCACT TGGGGTTGCG ACCATGCAGG CTGAAATGGC AACCAACTAT 24120

GGTTTGATTA TGGCAGGAGC TGCCCTTGCT GCTGTTCCAA TCGTCACAGT CTTCCTAGTC 24180

TTCCAAAAAT CCTTCACACA GGGTATTACT ATGGGAGCGG TCAAAGGATA ATACTCTGCG 24240

AAAATCTCTT CAAACTACGT CAGCTTCACC TTGCCATACT TAAGTATTGC CTGCGGTTAG 24300

CTTCCTAGTT TGTTCTTCAA TTTTCATTGA GTATAGGAAA ATCAATCTAT CAAGATACAG 24360

AAGTATATTT TATAGATTTA GAGAATATAG AGGTTATAAG TGTCTACAAA ATGGAGGGTA 24420

TGCAGTTACT TTATGAAGTT TTGTCAGACA CTTATAAACT TAAGAATGGT TTTAGTTAAC 24480

TATCAGAAAC GAAGGAAAGA GTATGATTTT TGACGATTTG AAAAACATCA CCTTTTACAA 24540

AGGGATTCAT CCTAATTTAG ACAAGGCTAT CGACTATCTC TACCAACATC GTAAGGATTC 24600

TTTCGAATTA GGAAAGTATG ATATTGATGG AGATAAAGTC TTTCTAGTTG TTCAGGAAAA 24660

TGTCCTCAAT CAAGCTGAAA ATGATCAATT TGAGTATCAT AAGAACTATG CAGATTTGCA 24720

TTTGCTGGTA GAAGGACATG AATATTCGAG CTACGGTTCA CGTATCAAAG ACGAGGCAGT 24780

AGCATTCGAC GAAGCGAGTG ACATTGGCTT TGTTCATTGT CATGAACACT ACCCACTCTT 24840

GTTGGGTTAT CACAATTTTG CGATTTTCTT CCCAGGTGAG CCACATCAGC CAAATGGTTA 24900

TGCAGGCATG GAAGAAAAGG TTCGAAAATA TCTCTTTAAA ATTTTGATTG ATTAAAAATA 24960

GGATGAATTG TTTTTTTGTA AAGCTTTGAT AATACTCTAC CATGAAATTG ATCTTTGTGA 25020

GGTAGAGAAA TGAGAATAAA ATATTTAAAA ATTGGTATCT TCTAAGTATG CTGCAAGAGC 25080

TAGTTTCTTA GATGGACAGG GGATTACAGT TGATGAGATG GCTTGGATAA TTAGGGGCAT 25140

TGTGAATGCA TTGATTGGTA GATACATAAA ATTAGGTACT TATGCGGCTA AGTATGGTAT 25200

TAGTATGGCA CGCTCGATCT TAAGTAGGGT AGCTGCAACT GCAGCAGCAA GAGTAGGATT 25260

ACTGACCAAG ATTTCTGGAT GGATTTTACG AGTAGCTGTG AATGTAGCTG ATGTATATGG 25320

TAATTTTGCC AACAATATTG CTGCAGCTTG GGATGCATAT GATAAAATTC CTAACAATGG 25380

TCGTATAAAC TTTTAAAATG CGAGAATGAA AGCACTTTGT ATTTTTTTAT TGAATATGTT 25440

AGCTTGGACA GTGCTTGCAA TGATAATTCG TGGAGGGCTA GATGGATTTG ATAGGCATAC 25500

TTGGAGTACT ATTTTAATTG CGTCGCTGTT CGGGGTATAT GATTATAAGC CCATAGATAA 25560

AAATAGAAAA AAGTCCAAAA GAAAAAATAG ATTTGTTCAT GGTAGGGACT TATGAAAGCT 25620

TTACTGACAA AAAAGAAAAC AGTTTACAAA GAAAAATGAT GGAGGAGCAA ACATGGCACA 25680

AAAAGGAGTA AGCCTTATCA AGGCAGCATT TGATACAGAT AACTTTCTCA TGCGTTTTAG 25740

TGAGAAGGTC TTGGACATCG TGACAGCCAA TCTTCTTTTT GTCGTCTCTT GTTTACCCAT 25800

CGTGACGATT GGAGTGGCTA AAATCAGCCT CTACGAGACC ATGTTCGAAG TTAAGAAGAG 25860

CAGACGGGTG CCTGTTTTTA AAATCTATCT AAGATCTTTC AAGCAAAATC TGAAACTAGG 25920

TCTTCAGCTG GGTTTAATGG AGTTAGGAAT TGTGTTTCTT ACCCTTTCAG ATCTCTATCT 25980

TTTCTGGGGT CAAACAGCTC TGCCCTTCCA ATTGCTGAAA GCCATTTGTT TAGGTATTCT 26040

GATTTTTCTT ACTATCGTGA TGCTGGCTAG TTACCCTATC GCGGCACGTT ATGACCTATC 26100

TTGGAAAGAA ATTCTTCAAA AAGGATTGAT GTTGGCTAGT TTTAACTTTC CTTGGTTCTT 26160

CCTCATGTTA GCCATTCTTG TCCTCATTGT GATGGTTCTT TATCTGTCCG CCTTCAGTCT 26220

ACTCTTAGGT GGCTCAGTCT TCCTACTTTT TGGGTTTGGA CTATTGGTCT TTATCCAGAC 26280

TGGATTGATG GAGAAAATTT TCGCAAAATA CCAATAGGAG CTTTATTTCT GAAACTACTT 26340

TCAAAGGCTC CAAACGCTAT TCTATAAGCG AGAAACTAAA ATCGG 26385

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2716 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CCTGCCCGCA TTGCCCTAGG CATTAAGTAA ACATATAAAA GCATGTGAGA GACTGTTGGA 60

AAAGCGAGGA AATTTCCCCT CTTTTCCTCT AGTCTCTCCT TTCTTTTGCT GATTTTATTC 120

AAAGAAAATG ATATAATAGT AGTTATGGAG AAAAAGAAAT TACGCATCAA TATGTTGAGT 180

TCAAGTGAGA AAGTAGCAGG ACAGGGAGTT TCAGGTGCTT ACCGTGAATT AGTTCGTCTT 240

CTTCACCGTG CTGCCAAGGA CCAATTGATT GTTACAGAAA ATCTTCCAAT CGAGGCAGAT 300

GTGACTCACT TTCATACGAT TGATTTTCCC TATTATTTAT CAACCTTCCA AAAGAAACGC 360

TCAGGGAGAA AGATTGGCTA TGTGCATTTC TTGCCAGCTA CACTTGAGGG AAGTTTGAAA 420

ATTCCATTTT TCTTAAAGGG AATTGTGAAA CGCTATGTAT TTTCTTTTTA CAACCGGATG 480

GAGCACTTGG TTGTGGTCAA TCCTATGTTT ATTGAGGATT TGGTAGCAGC TGGTATTCCA 540

CGTGAAAAAG TGACCTATAT TCCTAACTTT GTCAACAAGG AAAAATGGCA TCCTCTACCA 600

CAAGAAGAGG TAGTCAGACT GCGCACAGAT CTTGGTCTTA GTGACAATCA GTTTATCGTA 660

GTAGGTGCTG GGCAAGTTCA GAAACGTAAA GGGATTGATG ACTTTATCCG TCTGGCTGAG 720

GAATTGCCTC AGATTACCTT TATCTGGGCT GGTGGCTTCT CTTTTGGTGG TATGACAGAT 780

GGTTATGAAC ACTATAAGAA AATTATGGAA AATCCCCCTA AAAATTTGAT TTTTCCAGGC 840

ATTGTATCGC CAGAGCGGAT GCGCGAATTG TATGCTCTAG CGGATCTTTT CTTGTTGCCT 900

AGTTACAATG AGCTCTTTCC TATGACTATT TTAGAAGCTG CGAGTTGTGA GGCTCCTATT 960

ATGTTGCGTG ATTTAGATCT CTATAAGGTG ATTTTGGAGG GAAATTATCG GGCGACAGCG 1020

GGTAGAGAAG AGATGAAAGA GGCTATTTTG GAATATCAAG CAAATCCTGC TGTCTTAAAA 1080

GATCTCAAAG AAAAGGCTAA GAATATTTCC AGAGAGTATT CTGAAGAGCA TCTGTTACAA 1140

ATCTGGTTGG ACTTTTATGA GAAACAAGCC GCTTTAGGGA GAAAGTAAAA AGTGAGGTAA 1200

TCTATGCGAA TTGGTTTATT TACAGATACC TATTTTCCTC AGGTTTCTGG TGTTGCGACC 1260

AGTATTCGAA CCTTGAAAAC AGAACTTGAA AAGCAGGGAC ATGCTGTTTT TATCTTTACG 1320

ACGACAGATA AGGATGTCAA TCGCTACGAA GATTGGCAAA TTATCCGCAT TCCAAGTGTT 1380

CCTTTCTTTG CTTTTAAGGA TCGTCGCTTT GCCTACCGAG GTTTTAGCAA GGCACTTGAA 1440

ATTGCTAAAC AGTATCAGCT AGATATTATC CATACTCAGA CAGAATTTTC TCTTGGCCTG 1500

TTGGGGATTT GGATTGCGCG TGAATTGAAA ATTCCAGTCA TCCATACCTA TCACACCCAG 1560

TATGAAGACT ATGTCCATTA TATTGCTAAG GGGATGTTGA TCCGGCCGAG TATGGTCAAG 1620

TATCTGGTTA GAGGTTTCCT GCATGATGTG GATGGGGTTA TTTGCCCTAG TGAGATTGTC 1680

CGTGACTTGC TATCTGATTA TAAGGTCAAG GTTGAAAAAC GGGTCATTCC TACTGGGATT 1740

GAATTAGCCA AGTTTGAGCG TCCGGAAATC AAGCAGGAAA ATTTGAAAGA ACTGCGTAGT 1800

AAACTAGGGA TTCAAGATGG TGAAAAGACG TTGCTTAGTC TTTCGAGAAT CTCCTATGAA 1860

AAAAATATTC AAGCAGTTTT AGCAGCCTTT GCTGATGTTC TGAAAGAGGA AGACAAGGTT 1920

AAACTGGTAG TAGCTGGGGA TGGCCCTTAT CTGAATGACC TCAAAGAGCA AGCCCAGAAC 1980

CTAGAGATTC AAGACTCAGT CATCTTTACA GGGATGATTG CTCCTAGTGA GACGGCTCTT 2040

TACTATAAAG CGGCGGATTT CTTCATTTCG GCATCGACAA GCGAAACGCA AGGTTTGACC 2100

TACTTGGAAA GCTTAGCCAG TGGAACACCT GTCATTGCTC ACGGAAATCC TTATTTGAAC 2160

AACCTCATCA GTGATAAAAT GTTTGGAACC TTGTACTATG GAGAACATGA TTTGGCTGGT 2220

GCTATTTTGG AAGCCCTGAT TGCAACACCA GACATGAACG AGCATACCTT ATCAGAGAAA 2280

TTGTATGAGA TTTCAGCTGA GAACTTTGGG AAACGAGTGC ATGAGTTTTA TCTGGATGCC 2340

ATTATTTCAA ATAACTTCCA GAAAGATTTG GCTAAAGATG ATACGGTCAG TCAGCGTATC 2400

TTTAAGACAG TTTTGTATCT TCAGCAACAG GTGGTTGCTG TACCTGTAAA AGGATCTAGA 2460

CGCATGTTGA AGGCTTCAAA AACACAGTTG ATCAGTATGA GAGACTATTG GAAAGACCAT 2520

GAAGAATAGA AAGAGGAACA GCTATGAAAA AAACAATTAA TGAGAAGCGG TCGTGATAAA 2580

AAGATTGCGG GTGTTTGTGC TGGGGTGGCC CATTATCTGG ATATGGATCC GACTATCGTT 2640

CAAGTCATTT GGGGTGTTCT TACTTGCTGT TACGGAGCTG GAATTGTAGC TTACATTATT 2700

TTATGGATTA TCGCGA 2716

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13926 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTTTGGTTTT GCCTTATTCA AGACATGAGG GCCATCAGGA ATGATCTGAA ACTGCGAATC 60

TGTTAACAGT CTATGGAGAG CTTTCATAGA ACTAAGATTC GGTTTATCTT TGCTGCCACA 120

AATTAGTAAG GTTGGATAAG GGTAAGTTCC TGCTATATCC GTTAAATCAA GTGTCTTCAA 180

CTCCTCAGAA ACTCCGACCA TAAGAGTCTT GTCTGCTCCC TGTTTTTCAA ATACTCTTTT 240

GGGAAGTAGT TTAAAAATCA GCAATTGAAG ATAAAATAGG ATATTCCCTG CTAATTTAAG 300

CGGGCATCCT GACAGAATCA AAGCTCGAAG ATTTGGTAAA TCGTAACTGG AAAGTTCTAG 360

TGTCAGGGCA GCACCTAAGG ACAATCCAAT CAAAACAAAA GGTTCTGTCT CTTGAGCTAG 420

GTGCTGATAA ACTCGCTCTT TAGCTTGTTG ATAGTTACTA ACTCCAGAAG GAAATAACTC 480

GATAGCCTCA GAAGGATAAT CTGTCAGTAG ATTCCGAACT TCTTTCCAAG ACTCTGCTGA 540

CTGCCCTAAC CCATGCAAAA ATATTAATTT CATCTAGTTC TCCTCAAGGC TTAATTCATA 600

CAAGCCTCTC ACTGCATTAC AGCCGTAAAT AGCTTCTGCT TGGGTTAAAT CTGCCAAGGT 660

CAAGACTTTC TCTTCTACCT GTCCTGTTTC TAGCAAATGC TGACGGTAAA TTCCTGGCAA 720

GATTCCAAGT CGGATAGGCG GTGTGTAGAG TTTTCCAGCG ATTTTCAGAA CCAAATTTCC 780

TATAGAGGTT TCAAGCAGTT CTCCTGACTT ATTGTGGTAA ATCTTCTCTT GTTCTCCTAG 840

GCTCAAATGC GGTCGGTGAG TGGTTTTAAA GTAGGTAAAG GATTGATTCA AAGCAGCTTC 900

CTGAAGACAG ACTTGGGCCT GACAAAAGCT TGTACTGAGA GGGGTTAATA CTTGACGATT 960

GACTTCTATC TCTCCAGATT TGCTAAGGCT GATTCGCAAG CGGTAATCTC GATTAGCTTC 1020

ACAATCCTGA CACTCTTCCT CAATCTTGTG TCCCAAGTCT TCTGCATCAA AAGGAAAAGC 1080

AAAATAACGA CTAGCTTTTC TCAGCCTTTC CAGATGTTGT TCTTCAAACA TCAGTTGTTT 1140

TTGGCTGATT TTTCCAGTTG TAATTAATTG GAAGCGAGCT TGTTTACGAT AGAGAACTGC 1200

TGCCTTTTGA TGAACCTCTC GGTATTCAGA TTCCCATGTG CTATCCCAAG TAATCCCTCC 1260

GCCAACTCCA TAAATGGCTT GACCTTTGTG AAGTTGAATG GTACGAATGG CCACATTAAA 1320

AATCCGTCGT CCATTTGGAA GCAAGAGACC AATCGTTCCA CAGTAGACTC CACGCGGTTG 1380

AGGCTCCAAG TCCTTGATAA TCTCCATTGT CGCAATTTTC GGTGCACCCG TTATGGAACC 1440

ACAAGGAAAG AGTGAGCGGA AGATTTCAAC AAGGTCCACA TCCTCTCGCA ACTGACTCTT 1500

GATGGTCGAA GTCATCTGCC AAACAGTTGA ATACTGCTCT ACCTGACACA GACGCTCCAC 1560

GTGCTCGCTC CCAACTTCAG AAATACGGTT CATATCATTG CGCAAGAGGT CCACAATCAT 1620

CATATTTTCA GAGCGATTTT TGGGATCCTG TTCCAACCAA CTGGCCTGTT CAAGATCTTC 1680

TTGGTCAGTT ACCCCACGCT GAGTCGTCCC CTTCATTGGT CGTGTTGTCA ACTCGCGATC 1740

ATTTTGCTCA AAAAAGAGCT CTGGGCTCAT GGAAATCACT GTCATCTCGT CATGTTCCAC 1800

ATAGGCATTG TAGCCCGCCT CCTGCTCTAC CACCATACGA TTGTAGATGG CAAAAGGATT 1860

GGCATTTAAC TTTTGCTTAA GTTGGACGGT GTAGTTGACC TGATAGGTAT CTCCCTGCCG 1920

TAAATGATGG TGAATTTGGG CAATGGCCTT TTCATAGTCT GCTGCAGACG TTACTTCCTG 1980

CCAATTTGAG GGCAAATCAA TATCCTCATA AGTCAGAGGA ATAGGGGAAG TTTCTACGAT 2040

ATCATGAACA GTAAAGTAAA GCAGGTACTC TCCCAGTAGG GGATCCTTGT GAACTGCTAA 2100

TTTTTCCTCA AAAGCAGGTG CAGCCTCGTA GCTGACATAC CCCACCACAT AATAACCTTG 2160

CTCTTGGTAG CTTTCCACTT GTGCCAGCAA ATCTGCCACT TCTTCTACAT TTCTCGTTTT 2220

CAACTCTTTA ATAGGCTGGG TAAAGGTATA TCTCTCCCCC AAAGTCCTAA AATCAATCAC 2280

TGTTTTTCTA TGCATACCTT AAGTATAGCA TAAAATAAGA AAACCCTCAT CCGCAAAGCA 2340

GATGAGAGAT TTCAATTATT TAAAGATTGA AGTTTTAAAG CTATTTGTTT GTTGAAGAAG 2400

TTTCTTATAA ACAGCTTCTT TTAATTTAAC TGTATTATTC ATAGATACTG TTTTATTACC 2460

GTTTGCTTCT TGTTTAAGAG TTTCGGCATC TTTTTTAACA GCTTCTTTAA ACAATGTCAG 2520

TAAATCATCG TATGATGAAA CGGAAGAACC ATTTACTTCG AATGTTGTTA ATCCTTTCGT 2580

TGCTTTATCT TTAACTTCTT TGAAGTAAGC TTTTTTAAAT TCTTCAATAG TATTAAATGT 2640

ATTGTTAGAT ATTTTCTTGA TAATATATTC ATCACTTAGA ACAGACTCAC CATCTGTTTT 2700

AGATTGTTGT TTATATTTAT TTGAAGCATA ACCTAAGAAC CCATTTTCGT ATCCGTAGTA 2760

ACCCCATAAT CTAAAAGCAT TATGTTTGAA TGAAACAGCT CCAGGAGCAC CTTTACTAGT 2820

ATTACCTCCG TAGATACCGG TCATCATTCT AACACCTACA TAAGGTGATT GATCGTTATA 2880

GCTAATTGCT TCGGGTTTAT AGATACCATT ACCTGGATTG CGATTAGTCA TTAATTGTTG 2940

ATCAACTAAA TCATTAACAG ATTGAATATT TAATTCATTT TTCTCTTCTT GACTTAGATT 3000

TCGAATTTTA TCCCATTGAT TTAATTTATT GTTATCACGG TATTCTCTAT CTATTTTTTT 3060

GAACCATGCA CTATTTAAAT CTTTATTTTG TTGAGAAATC ACAGATTCAG CCTCAATTTC 3120

ATCAAGAAGA GTTAAAGTGT CATTATAACC CTTCATATAT CTATTAATAT CTTCTCGTGT 3180

TTTTAGAGTT TTTGGATCTG TAATATACCA CTGATTCCCA TCATTTTTGC GTTTAAATAC 3240

CATATTAATA CCTAAAGAAC CAAACTCATC AAATCCACTA CCAGTAACAG GAGTTTGTAG 3300

CATACCCTGA GCATATGCTT CAGCATCAGT ACCTTCACGG TGTCCAAAGC CACCTAAGTA 3360

AATCGCACGG TCGTTGACGT GTGTTGTTTC ATGTGTGTAA ACTGAAATAC CGTATTCACC 3420

AACCATTTCT AAATGAACAT ATTTTACATC AGTTCTAATA TCATCAGAGT TAGGATATAT 3480

AGCAGCATAA GCTCCTGTTC CATTATAATT ATAATACTTA TCCATAGGAC CAAAGAATTC 3540

TCTAAGAGGA GTATATACTT TGTCGGTATT ATAGCGGCCA TATTTTTCAA CCCATCCACC 3600

AGGAGCGTTA TAACCTTCCC AAATAGGAAT AACAGCATCT CTTAGTAGTC GTTGTTTAAC 3660

GTTATCAGAC GCTAGACGAT ACCAGAAATC ATAATAGTTT CTATAACCAT CTGCAGCTTT 3720

GTTAACGATA TCTTTAATAT CTTCTAATGA TTTTTTACCT AATCGCTCTG CACTACCAAA 3780

GGCAATTGCA TTATAATTTG AAATTAAATA AAGATGTGCT TTATCAATAT TCAGTAGTGG 3840

GAGTATAGTA TTTCTAAGGT GACTTCGTTT TAAATTATCG AATGCACGAT GTTTAGAATT 3900

TTTAATTTCT TCGACCTCAG AAGCGCGTTC TGCGATGTAG ACATGGTCTT CTGTAGCATC 3960

AATAAACCAA TCGTTCATAT TGTCTATATT TGTGAACAAT TGTCTATTAT AATTTAAAAA 4020

TGCATCTAAA TTACCTGATT TAGTATATTT AGCCAATACT TGACCGAATG CGTCGAATGT 4080

ACGTGAACCT TTAATGTTGT TCTCTTTAGA ACCGATTTCA ATTAATCTGT CTAATACGCT 4140

AACTTTTTCA CCATAGAAAT CTGGTTTGAA TAGCATTAAT TCTTTAATAT TAACATCACC 4200

AAATTTAACT CCATAGTAAC GATTTAGGTA AGTTAAACCT AGTAATAAAG CTGCTTTGTT 4260

TTTCTCGACT TTATCACGAA TCATTTGACG AGCAGCTGGA GAATCATTTA GTTGATGTTC 4320

TTCGTTTTGA ACTAATTTTG TGATTAGGTT TGTTAAGTTT TCTTTAACAT CTGTGAAGCT 4380

TTCTTCTAAA TATAAATCTT TGATTGCATT AACTCTATAG TCACCTAATC GATTTAGATG 4440

CTGATACATC GTTTGAGACT GAAGCTCTAC TGATTCTAAA ATAGATTTTA TATCATTAAC 4500

AAGAGTAGTG TTATCTTTTT GAACGATATT AGGTGTATAT TTAATTCCTA AGTCAGTTAT 4560

AGTATATTCT TTTACATTAC TTAAACCTTC ACTGCTAGAA GACAAGTTAA AGTAATCTTT 4620

TGTACCGTCC GCATAGTGAA CAATAATTTT ATTAGCTTCA TCTAGGTTTG TGATAAACTC 4680

ATTGTTGTTC ATCGCGGTAA CAGAAAGAAC TTCTTTAGTA TTTAGATGGT GTTCTTTATT 4740

TAATTTATTA CCTTGATATA CAATATAATC TTTATTGTAG AATGGTATTA ATTTTTCAAG 4800

ATTTTTATAG GCTTGGTTAT ATTCAGCGTT ATAATCTTGA ATACTAGAAT AGGCTTTTTC 4860

TTCATTAAGT TTTGCAAGAG GAGATAGATC ACTTTCTAAT TTATCAGCAG TAATATTGAA 4920

AGTAGTAACT TTAGCATCAG CTTGTTCTTT AGTTAATTTA GTAAATGTTT TAGATTTCCT 4980

AAATGATCTA TTACCTGACG AATATCCCTC TACCGCATAT AAATCTTTTA TATGAGCACT 5040

AGCATAATCA GAATCATCAA CGTCGTTAGA GCCGAATAAC TCCTCTCCAC GGATAATCTT 5100

AGCATAGCTG ACAGAATTAC TTACCGTACC TACAGGCCAA GTCTTACTTG CTATTGCTCC 5160

AACTTCTACT GGATTTGAAA CATCTATTTT ACCTTTTACA ACCGACTCAG TTAGGAGAGC 5220

TTTTGTACCA ATAAGATGGT CTAGAGTTAA TCCATAATCT ACTTTAGGAA CTAACAAGCT 5280

GGCGCGTGTT TTGTTTCCTG TAATAGTAGC ATCAACATAT GCTTTTCTAA CAATTCCTCT 5340

ATAGTTTGTA CCTGCAATTC CCCCTGTATG AGAGCCATTT CCACTTGTAG AGTGTAGTTT 5400

GCCAAAGAAA GCAACATTTT CAATACGAGT TCCATCATTC ATATTATTTA CAAATCCAGC 5460

AACATTATTA CGACCTGAAA GTGTGCCTGT AATTTTGACA TTTGTAATAA CTGAAGAACC 5520

TTTCATAGTA TTGGCTAATG ATGCAATATT ATCTTGACCA GAACGTTCTA TCTCTACATT 5580

TTCAAAATTC ACATTATTTA TCGTTGCGTT TGTTATCACA TTAAATAATG GATGTTCCAA 5640

TTCAGTAATA GCAAATTGTT TTCCTTCAGA ACTTAAAAGT TTTCCTGTGA ATTCTTTAGT 5700

GATATATGAT TTTCCATTAG GAACAACATT TCTAGCGCTC ATTGATTGTC CCAGACGATA 5760

TTCTTTTGAA GGATCGTTTT GAATAGCTTC CACTAATTCT TTGAAATTAT AATATACATT 5820

ATCTTCGTGG ACTTTAGGTT TTTCAATATA GTGAACGTAT TCTTCTTCAA ATTTATTATC 5880

AGCAGTTCTA GAGACTAAAT TGTCTGCGAT TGCTGTAACT TTATATACAG GTGTTCCGTT 5940

AACCGTAGTT TCTTCTATAT TTTTAACAGC TAGTAATGTA GTTTTCTGAT TATTTGAAGT 6000

TATTTTTAAA TAATAATTGC TCTTATCATC AGGAATAGTT GTTATCAGTG ATTCATTAGT 6060

TTCTTTTCCA TTTTCGTATT TGATTAAATC TGTACGTTTA ATATTTTTAA GCTCAACTTT 6120

TTTAAGATCT AATTGAATAT TTTGATTTTC TAGAGTTTCA GTTTCTTCAC CGTTACCTCT 6180

GTCGTAAATC ATAGTTGTAG ATAGGGTGTA TTCTTTGTAG TACTCTAGGT TCTTAAATGC 6240

AGCGCTTATA GTTTCTGTTG TTACCTTGTC ATCTGTAAGG ACTACAGTAT TAATAACTTC 6300

TTCTCCTTTT TTCAATTCAG CTGTGATTGA TTTGATTTTT GTTTTGTTTT GATTTTCTAG 6360

AGTATACTTA GCAACAGCTT CACGTTCCAA TATTTTCTTA TCGGTACTAG TCAATGTTAA 6420

TATTGGCTTT TCAGATAATT CAACCAATTT TTCAATAGTT GCAGTTAATT TTTCAACAGC 6480

TTCGTTAACT TCACTTTGTT TAGCATCTGT ATTAGCTGCA ACTTTTTCAG CCTTTGTAAC 6540

TTCAGTTTGG AGGTTTTGCC AACTTCTATC ACTGTAATGT TCTTTTACCT TTGTTTTTGC 6600

ATCTGCAATC GTATTGTTTA ATTCAGTTTT ATCAACGTTT AGAGCGTCAA TAGCCGTTTT 6660

AAGTTTATTT GTCTCGCTAT TTACCTCAGG CTGTTTTACA GGCTCTGAAG CATAGACACC 6720

TTTTGCAGTT TCTAAAACAG GTCCAAGAGC ATTGTAACTT GCTGTAGAAT AATCAGTAGG 6780

AGAAACTGAA CTAGCTTTAT CAATTTGATT ATTTAACTCA CTTTTATCAA CTGGTTCTTT 6840

AGTACCAATA CCCTTTATTT TATCTTCTGG TTTCGGTGTT TCCTCTACAG CCTTCTCTTC 6900

TTCAGGAACT TCTGGTTGCT TTTCTGGCTC AACTGGTGCC GTTGGTGCCT GTTCGTCTTC 6960

TCTTGGCGCG ACTGGTTCAC CTGCTTGTTC AACTTTTGGT TCCTCTGTTG GTTCTGTTTG 7020

TTTTTCTACA GCAGGCGTTT CAACTTTTGG TTGTTCAATA GATTGATTAA CAGTCTCCTC 7080

TTTTGGTTCT ACAGTTTCTT CAGCCTTGGT ATCTGGAGTT GACTCTTCTT GTTTCGGTGT 7140

TTCCTCTACA GCCTTCTCTT CTTCAGGAGC TTCTGGTTGC TTTTCTGGCT CGACTGGTGC 7200

CTTTTCGTCT TCTCTTGGCG CGACTGGTTC ACCTGCTTGT TCAACTTTTG ATTCCTCAGC 7260

TGGTTTGTCT GATGGTTGAC TTTCTGGCTT AACTGCTACT TTTTCCTCTG GTTTTGACTC 7320

AACTTCTCCA CCTACTTCTT CAACTGGAGC TGGTTCTGCT GAATCTTCTT TCCCCTCTTC 7380

TACTTTAGGA AGGGTGTCGT CAGTAGGTTT TACCTCCGAT TTTGGTTCTT CCTTTGGACT 7440

TTCTTCTGTT TTAGGTGCTT CTTCTTTTGG AGCTTCCTCT GTCTCTACTA CTTGGTTTTC 7500

TGTCCTAGCT TGCTCCTGAT TTGTTATTGA TTGAGGAGTC TCAACTTCGA CCACAGTCAC 7560

CTCTCCAGGT TTTGCTGAGG TTTCTTCTAA AACAGTGTCC AAGCCAAGCG TTTTGAGGAT 7620

GTCACCTGAT AGATAACCAA CATAGCGATA GCCCTCCATT TCAACAACAC CCTCTCGACT 7680

AGCCAGCGCT AGGGTCGCAA CTGGGTCTAC AGCCCCTGCA CTAGGAAGAA CTACCAATCC 7740

CATAGCTCCA ACTAGAAAGA CGCTAGCAAT TTTCTTTCTC TTGTAGATTA AAAGCAAGCT 7800

CCCAACAGTC AGCAAACCAA AAGCTGTCAA AACAGATGCT TCTGTCCCTG TTTGAGGCAA 7860

CTGATCTTTT TGATACACCA AACCATATAC AACTTCATTC CTGTCAGGCT TTCCTGTCTG 7920

AATTAAATCT TTAGCTTCTT GTGAAATAAT CTCTTTATTT ACATAGTGAT AGGTGGCTGC 7980

GTCCACTACA GAAGGAGCCA TCAAAAGGCT TCCAAGAAAT ACAGAGCCTA CAACTCCCTT 8040

AATCTTACGA ATTGAAAAAC GGTCTTTTTT AAACACTTTT ATCTCCTTTA TTCATTCTCA 8100

AAACTTCCTA ATAGCATCTT GCGGATAGTG CGCACGCGCA CCTCCGATTA ATTTTGGACG 8160

ACTAGCCAGT GCCGTTACAT GGGCATGACC AATCTCTCTC AAAATAGGGC GAATCGGAAC 8220

CTGAACATGC TTGACATGCA TGCCAATTGC AGTGTCTCCG ATATCCAATC CAGCATGAGC 8280

CTTGATAAAT TCAACCTCAA CTGGATCCTG CATAAACTTA AAGGCTGCCA ACTGCCCCGA 8340

ACCTCCTGCA TGAAGAGTAG GATGGACACT GACAATTTCC AGACCAAACT GCTCTGCCAC 8400

CTGACGTTCA ACAACGAGAG CCCGATTGAC ATGCTCACAA CCTTGAACTG CTAAATGGAT 8460

ACCTCTACTA CCTAGAATAT CCAAGATAGT CTCCACTATC AGCTCACCAA TCTCTTGACT 8520

GGATTCTTTC CCAATATGAC CACCTAGCAC CTCACTAGAA GATAGACCTA AAACAAAAAG 8580

GGCCCCCTGC TTCAAATTGG TCTTTTCTAA AACATCTTCC ACTACCTGAC GTGTTTCTCT 8640

TTGAATCTGT GTCTCGTTCA TCTCTGTTAC CTCTGTTGTC ACTCTTCTAT CATACCGTTT 8700

TTTCTTGTTT TTAGCAAGAT AGACAACCTA GAAAGTTTGC CCAATTACGC ATAAAACTCC 8760

CAGAATTGAC TGGGAGTTAG CTAGTTTCTA TTCTATTTAT ATATATTTCA ACTTTCGTCC 8820

CTTTTTGGGG TCTAGAATCA ATCTTCATAT GGTAATTGGC TCCAAAATGA AGTTTGAGCC 8880

GTTGATCGAC ATTTTGAAGA CCAACTCCCC CACGTTTGAG TTGACTTTGA CTACTATCAC 8940

CAGCATCTTG GAAGCCAACG CCATCATCCT CAATACGGAT GACCAATCCC GAATCCTGTT 9000

TCTGGACAGA AAGTTTAATA TGGCCCTGAC CTTCCTTTTC CTTAATGCCA TGGTAAAGAG 9060

CATTTTCTAC AAGGGGTTGT AGGACCAGCT TGGGTAAGAC TAAATTATCA AAGGCAACAT 9120

TTTCATTAAT TTCGTATTCC AGCTTATCTC CATAGCGTTG TTTCTGGATA AAGAGATACT 9180

GGCGGACATG ATTGATTTCG TCAGAGAGAC AAATCAAGTC CTTGCCTTGA TTGAGCGCCA 9240

AGCGGAAATA GGTTGCCAAG GACTTGGTCA CCTGCACCAC TCGCTGACTA TCATGAAATT 9300

CAGCCATCCA GATGATGGTG TCCAAAGTGT TATAGAGGAA ATGTGGATTA ATCTGGCTCG 9360

AAAGGGCTTG AAGTTGGTAC TGACGGGTCG TTTCTTCCTG GCTACGAATA GCTACCATCA 9420

ACTGATCAAT CTGATCCAAC ATAGCATTAA ATTGGCGAGT TACTTCTCTC AGTTCATAGG 9480

CACCAACTTC CTTGGCACGA AGATTTTGAG CACCAGAAGC AATTTCCAAC ATGGTTTCTC 9540

TCAAATCCTT CAAAGGAGCA ATCCAGCGTT TAAGACTGAA CCACACTAAG CAGAGACAGA 9600

CAAGAAGAGA TGTGACACTG GCCCCAAGCA AGGTCCACAA GAGCTGACTC CGAACCTGGT 9660

CTAACTTTTC CAATGATGAC ACGCCAAGCA CCGTCCAATC AGTTCCTGCA ATCTTCTCTT 9720

GACTGACGTA GGATTTGTGA CCAGGAGTAT AACCCTGACC TGTATCGATG TAGGGTTTCA 9780

TAGCCTCCAT TTTGCTAGAC GAACTATAAA CTGTGTGTTG AGGATGGTAG ACAAATTCAT 9840

GGTTTTCATT GATAATGAAG GCAAAGCCCT GCTGCCCCAA CTGGAGTTGA TTGAGATAGG 9900

CTTCCAGAGT TTCATAAGAA ATATCCAAAC GAAGCACACC AAGATTGGCT CCCTTTGCAT 9960

CAACAAGTTC TTGAGTGACA GAAATGACCC ACTGACTATC TGATTTACGA GCTGGAGTCA 10020

AAACAGGCAT AGCTCCCTGA TGAATGGCCT TTTGGTACCA ATCCTCAGCC ATCATATCAG 10080

AGGAAGTTTT CATCTGCACA CTGTCATCTG TAGAAATGAC CTGACCAGAT TTGGTCACCA 10140

GCACAACAGT TTTCAAGTCC TTATCTGACT TCAAGATGGT CAAAAACAAA TCTCGGATTC 10200

CCTCGACCTT GTCTTGACTG GGATTCTCAG CATAGGCCAG AACATCCGTC TGCTGGGTCA 10260

AACCAGTCGA GGTGGTTTCT AGTTTTTTGA TATAAGACTG AATAAAGTGG CTAGTCTGGC 10320

TGATGGTCGT TTGGCTGTTG CCCTCAATGG TGGCCTCAAT GGCTGAAGAA CTTGATTGAT 10380

AGTAGAAAGT TCCAACCAGA GCTAGGAGAA TGAGAAAGAC CAGAAAGATG GAAATAACCA 10440

TTCTAACTAA AAGAGAAGAA CGCTTCATCG GTCTTCTCCC TTCTTAAACT GACGAGGTGT 10500

CACACCTGCA ATCTGCTTAA AACGTTGGGT AAAATAGTTC ATATCTTCAA AACCAACCTT 10560

CTCTGCGATC TCATAAATCT TCAGATCTGT AGTTAAAAGC AAGAGCTTGG CTTGTTTAAC 10620

ACGTTCTCTC ACCAGATAAT CCTGAAAAGG CAAGCCCAAC TCTTTCTTAA TCAAGGAACT 10680

CAGATAGGTC GGACTAAAAC CTAAGTCACT GGCTAAAGAC TTTAAACTAA ATTGGCTATC 10740

AGCCAGATGA GACTGGATTT TCTGGGCCAT GTTTCCTTCA AACCTATTAG TCAATAAATC 10800

TTGTAACTGC TCTTCTTTCT CTTCCTTGTC TAGTTTTTGT TTGATTTTCC CCAACATTTC 10860

CTCAATATCC TGACGAGAAA AGGGTTTGAG CAGGTAGTCG TCCACACCTA GTTTGACAGC 10920

AGACAAGGCA TAATCAAAAT CATCGTAACC TGTTAAAAAG ACCAAATGAA CCTGAGGATA 10980

GGTTTCTCGT ACCAGACTGG CCAACTGGAT GCCATTTAGA TGAGGCATGT TGATATCGGT 11040

TAAAATGATA TCTGGCACCT GCTTTTGGAT CAATTCCCAA GCCTGCCTTC CATTTTCAGC 11100

CTGACCGATG ATTTCCATAT CGTAGGCTGC TACATTGACC AGTTTAGTCA AACCTTGTCT 11160

TACCAGATAT TCATCTTCTA CGATTAAGAT TGTGTAGGTC ATGCTCTGCT CCTTTACCAC 11220

TTACTAGTAT CAGTATAGCA AAATTCTCCT CTAACTGCTT AGGAAAGACC TCTTATACTC 11280

AATAAAAATC AAAAAGTAAA CTAGGAAGAT AGCCACAGGT TTCTCAAAGT ACCGCTTTGA 11340

GGTTGTAAAT AAAACTGACG AAGTCGACTC AAAGTATAGC TTTGAGGTTG TAGATAAAAC 11400

TGACGAAGTC GATAACCCTA CATACGGTAA GGCGACGCTG ACGTGGTTTG AAGAGATTTT 11460

CGAAGAGTAT TAATCAACAT AATCTAGTAA ATAAGCGTAc CTTTTTCTTC CATTTGGTCT 11520

TTGGGAATAA AGCGGATAGA GAGGCTATTG ATACAGTAAC GTAAGCCGCC CTTGTCCTGT 11580

GGACCATCCG TAAAGACATG CCCAAGGTGA GAATCTCCTA CTCGGCTCCG CACTTCCATA 11640

CGCGTCATAT TGTAGGACTT ATCTTCCTTG TAGGTGACAA CATCTGGACT GATGGGTTGG 11700

GTAAAACTAG GCCAGCCACA ACCAGACTCA AATTTGTCTT TTCATGAAAA GAGAGGTTCC 11760

CCAGTTGCTA TATCCACATA GATACCGGAT TCAAATTTAT CCCAGTAACG GTTTGAGAAA 11820

GCTCGTTCTG TTTGATTTTC CTGGGTAACT GCATACTCCT CAGGTGACAG GGTCTTTTTC 11880

AATTCCTCAT CACTTGGTTT TGGATATTTG CTGGCATCAA TGACAGGATA GGCCGCCTGA 11940

TTAACATTGA TATGGCAGTA GCCATTTGGA TTTTTCTTGA GATAGTCTTG ATGGTAATCC 12000

TCAGCCACCA CAAAATTCTT CAAGTTTTCC TTTTCAACTG CTAGAGGTTG ATCGTATTTC 12060

TTAGCCACCT CATCAAAGAC TTGGTTAATC ACTTCCAAAT CCTTGTCATC TGTGTAATAA 12120

ACACCAGTAC GGTACTGGGT CCCCACATCA TTTCCTTGTT TATTTTTGCT GGTTGGATTG 12180

ATAATGCGGA AATAGTGAAG CAGGATTTCC TTGAGAGAAA TTTGCTTGGC ATCATAGGTG 12240

ACATGGACGG TTTCTGCATG ACCTGTTTGG TTAATCAATT CGTACTTGGT TGTTTCTCCT 12300

CTACCATTTG CATAGCCTGA AACGGCATCC GTCACCCCGG GAACACGTGA GAAATATTCC 12360

TCCACTCCCC AGAAACAACC TCCAGCTAGA TAAATTTCGT GCAAGTCTGC GTCTTTACTA 12420

ATTTCTGTTT TTTTCACTGC TTTTCCTCCT TGGCTAACTG CCGCCTTTTC AATTTGCGAG 12480

GCATCTGTCT GCCCTGCATT TCGTATCAAT AGAACATAGA AACCGGTTAT GGCTAGAAAA 12540

AATACTCCTA GCAACAAGAA GATTTTTAAC TTATCATTCA TAAGACGCCT CCTAGGCTAA 12600

TTCCTTCAAA GTTTGCAAAA TTGCATCTTT TTCCATGAAT CCTGGATGTG TTTTGACCAG 12660

CTTGCCTTCT TTGTCTATAA AGGCTTGGGT TGGGTAAGAA CGGACACCAT AAGTTTCCAA 12720

AAGTTTGCCT GATGGGTCAA CTAGGACTGG GAGATTTTTA TAATCCAATC CCTTATACCA 12780

ATTCTTAAAG TCCGCTTCAG ATTGCTCTCC CTTATGTCCT GGTGACACTA CTGTCAAGAC 12840

CACATAGTCA TCACCAGCTT CTTTAGCAAT CTCATCCGTA TCTGGAAGAC TAGCCAGACA 12900

GATGGAACAC CAAGAAGCCC AGAATTTGAG ATAGACTTTC TTGCCCTTGT AATCAGATAA 12960

ACGGTAGGTC TTGCCATCTA CTCCCATCAA TTCAAAATCA GCCACCTCTT TCCCTTTAGC 13020

TGCGCTTGTT TTACTAGCTG TCTGCTCCGT CTTCATTTCA TCTTTCGTTT GGTGTTCACT 13080

AGTCACGGAC TTGCCTGAAC AAGCCGTCAA ACAAAGGAGC GAACCTGCTC CAAGAACACA 13140

TGTTTGCCAT TTTTTCATAT TGATATTCCT TTCCATTTTA TTCAAATAAT TGACTTAAAA 13200

TTGAAGCATT TCCAAACAGA ACCAAGAAGC CCATCACAAT AATGAGAAAA CCACCCACTT 13260

TTTTGAGGAT TCCGAGATAG GGATGAAGTT TTCGGAAATG TTTCAAAACA TAACTAGAGG 13320

TCAGAGCTAG AAGCAAGAAT GGTAGCGCCA AGCCCAGCGT ATACACCAAC ATGAGACCAG 13380

CTCCCTGCCA AGCTCCTGAA CCACCTGAAG CCGCCAAGGC CAAAACAGAC CCCAGAACCG 13440

GCCCCACGCA AGGCGTCCAA GCAAAACTAA AGGTCAAGCC CAATAAAAAT GCCTGACTAT 13500

AGCCCTTACC ATTTTGCCCC TGTCCTTGCA GTTGTAGCCT CTTTTCCTTA TAAAGCCCCT 13560

TAAAGTGTAG AATCTCCATT TGGTGCAAAC CAAGAAGGAT AATAATTGCC CCAGTAAGAT 13620

ATTGGAACCA AGAAGCATAA AGCAAATCGC CTAAAAAACC AGCTCCATAG CCCAACAAAA 13680

TAAATATAAA GGAAATTCCT GCTATAAAGG CCAGAGTTCG TAATAAACTA GTAACTGAGA 13740

TTGAAAATTT GCCGCTAGAA GCCTGAGCAC CATCCTTATC ATCTAGTAAC ACTCCTGTAT 13800

AGACCGGTAA CAAAGGTAAG ATACAAGGAG AAAAGAAGGA TAGAATCCCT GCCAAAAAGA 13860

CACTTAGAAA AAAGAAAATA TGACCCATAA AGTTCCTCCT ATCATTTTAT TGATAGATTT 13920

ATTATA 13926

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20199 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CCCAGCAGAA AAATGGCATT TGGAGATAAT GGAAATCGTA AAAAAACTAT GTTTGAGAAA 60

ATAACCTTGT TTATCGTGAT TATCATGCTA GTAGCAAGTT TATTGGGAAT TTTTGCAACT 120

GCAATTGGTG CCCTCAGTAA TCTATAAAAT AGATTCAAGA AAATTTAGTG ACTGGGATTT 180

CCCAGCCCTT TTTTAAAGTG AGAAGAAATA ATGAGTATGT TTTTAGATAC AGCTAAGATT 240

AAGGTCAAGG CTGGTAATGG TGGCGATGGT ATGGTTGCCT TTCGTCGTGA AAAATATGTC 300

CCTAATGGAG GCCCTTGGGG TGGTGATGGT GGTCGTGGAG GCAATGTGGT CTTCGTTGTA 360

GACGAAGGAC TACGTACCTT GATGGATTTC CGCTACAATC GTCATTTCAA GGCTGATTCT 420

GGTGAAAAAG GGATGACCAA AGGGATGCAT GGTCGTGGTG CTGAGGACCT TAGAGTTCGA 480

GTACCACAAG GTACGACTGT TCGTGATGCG GAGACTGGCA AGGTTTTAAC AGATTTGATT 540

GAACATGGGC AAGAATTTAT CGTTGCCCAC GGTGGTCGTG GTGGACGTGG AAATATTCGT 600

TTCGCGACAC CAAAAAATCC TGCACCGGAA ATCTCTGAAA ATGGAGAACC AGGTCAGGAA 660

CGTGAGTTAC AATTGGAACT AAAAATCTTG GCAGATGTCG GTTTAGTAGG ATTCCCATCT 720

GTAGGGAAGT CAACACTTTT AAGTGTTATT ACCTCAGCTA AGCCTAAAAT TGGTGCCTAC 780

CACTTTACCA CTATTGTACC AAATTTAGGT ATGGTTCGCA CCCAATCAGG TGAATCCTTT 840

GCAGTAGCCG ACTTGCCAGG TTTGATTGAA GGGGCTAGTC AAGGTGTTGG TTTGGGAACT 900

CAGTTCCTCC GTCACATCGA GCGTACACGT GTTATCCTTC ACATCATTGA TATGTCAGCT 960

AGCGAGGGCC GTGATCCATA TGAGGACTAC CTAGCTATCA ATAAAGAGCT GGAGTCTTAC 1020

AATCTTCGCC TCATGGAGCG TCCACAGATT ATTGTAGCTA ATAAGATGGA CATGCCTGAG 1080

AGTCAGGAAA ATCTTGAAGA CTTTAAGAAA AAATTGGCTG AAAATTATGA TGAATTTGAA 1140

GAGTTACCAG CTATCTTCCC AATTTCTGGA TTGACCAAGC AAGGTCTGGC AACACTTTTA 1200

GATGCTACAG CTGAATTGTT AGACAAGACA CCAGAATTTT TGCTCTACGA CGAGTCCGAT 1260

ATGGAAGAAG AAGCTTACTA TGGATTTGAC GAAGAAGAAA AAGCCTTTGA AATTAGTCGT 1320

GATGACGATG CGACATGGGT ACTTTCTGGT GAAAAACTCA TGAAACTCTT TAATATGACC 1380

AACTTTGATC GTGATGAATC TGTCATGAAA TTTGCCCGTC AGCTTCGTGG TATGGGGGTT 1440

GATGAAGCCC TTCGTGCGCG TGGAGCTAAA GATGGGGATT TGGTCCGCAT TGGTAAATTT 1500

GAGTTTGAAT TTGTAGACTA GGAGACTGGT ATGGGAGATA AACCGATATC TTTCCGAGAT 1560

GCGGATGGTA ATTTTGTTTC CGCCGCAGAC GTTTGGAATG AAAAGAAATT GGAAGAACTA 1620

TTTAATCGTC TCAATCCAAA TCGTGCCTTG AGATTGGCAC GAACTAAAAA GGAAAATCCA 1680

TCTCAGTAAA GAAGCTAAAA AATCCCGTGC CTCATCAGAC ACGGGATTTT GTGGTACGAC 1740

AGGCATGTAT AGCAAACTGA ATCTGGAATA GCACAGCATA TCTTCTAAAA TATAGTAAAA 1800

TGAAATGAGA ACAGGACAAA TCGATCAGGA CAGTAAAATC GATTTCTAAC AATGTTTTAT 1860

AAGCAGAGAT GTACTATTCT AGTTTCAATC AACTATATTG TTATAAATTG ATTTGAATTT 1920

CAAAATTAAA TTGTTTGATT CTTATTTCAA TTTGTTATAG TATATCTGAT GTCAAAGTTC 1980

TCGGCGAGTC AAATAGCGAT TCCCAAGCCT GACTATCGTG AGGTAGCGGA TTAAAATGGT 2040

CTGGGGATAG ACCGTTTTAA GTCTGACGCT GGAAATAAGA ATTGTCAGAA GAAGGGATAG 2100

CGAAATCGTG GCTCTACGAA CAGGAACGTG ATAATAAGGC GTATATAGCG GATAAGAGGG 2160

CATCAAACTC TAAAGTCCAA AAAGGTAGTC GTAACCTATA TGCGTAAATC ACGAGAGTAA 2220

TTGAATTCGT ACTAAGATTT TCTATTTTCA CTGTAACCTT TTAACGCCCT TATATCTTGT 2280

ATACACGAGG AAAGATGTAC GACTTATCCC GTGAGGTCTA TCACTATAAA GAGAAAACGA 2340

CAGATAGAAG TGATCCTGAG TCACGGTTAT CTGTCTGATA GGACGGTATG TATAAAACGC 2400

TTCTGTGAAC TGAGAGAAGG GGGAGAAGTT CTTGCTAAAA TTTAGTTGAA CAGCCGTATT 2460

CCGATACTTA GATAAGAGAT CTAGTCTTAG CTCCTACTCA GTTTTAGGGG ATAAAAAAGG 2520

GGCAATAGCG ATTCGAGAAA GATTATACTC TTCGAAAATC TCTTCAAATC ACGTCAATAT 2580

CGCCTTGTCG TATGTGTAGG ATACTGACTA CGTCAGTTCC ATCTACAACC TCAAAACAGT 2640

GTTTTGAGCA ACCTGCGGCT AGTTTCCTAG TTTGATCTTT GATTTTCATT GAGTATTAGT 2700

AATTCAGTTA CTAACTCGTC AACTCTGATT TATCCAATAA AATTGAAAAG GATGGAAAAA 2760

AGGATAAATT TATGATATAC TTTATTTTGA AGACCTTATT AGAAATCTTG AAAGAGTATT 2820

GAAAACTTAG AATGAGAAAA ATTGTTATCA ATGGTGGATT ACCACTGCAA GGTGAAATCA 2880

CTATTAGTGG TGCTAAAAAT AGTGTCGTTG CCTTAATTCC AGCTATTATC TTGGCTGATG 2940

ATGTGGTGAC TTTGGATTGC GTTCCAGATA TTTCGGATGT AGCCAGTCTT GTCGAAATCA 3000

TGGAATTGAT GGGAGCTACT GTTAAGCGTT ATGACGATGT ATTGGAGATT GACCCAAGAG 3060

GTGTTCAAAA TATTCCAATG CCTTATGGTA AAATTAACAG TCTTCGTGCA TCTTACTATT 3120

TTTATGGGAG CCTCTTAGGC CGTTTTGGTG AAGCGACAGT TGGTCTACCG GGAGGATGTG 3180

ATCTTGGTCC TCGTCCGATT GACTTACACC TTAAGGCGTT TGAAGCTATG GGTGCCACTG 3240

CTAGCTACGA GGGAGATAAC ATGAAGTTAT CTGCTAAAGA TACAGGACTT CATGGTGCAA 3300

GTATTTACAT GGATACGGTT AGTGTGGGAG CAACGATTAA TACGATGATT GCTGCGGTTA 3360

AAGCAAATGG TCGTACTATT ATTGAAAATG CAGCCCGTGA ACCTGAGATT ATTGATGTAG 3420

CTACTCTCTT GAATAATATG GGTGCCCATA TCCGTGGGGC AGGAACTAAT ATCATCATTA 3480

TTGATGGTGT TGAAAGATTA CATGGGACAC GTCATCAGGT GATTCCAGAC CGCATTGAAG 3540

CTGGAACATA TATATCTTTA GCTGCTGCAG TTGGTAAAGG AATTCGTATA AATAATGTTC 3600

TTTACGAACA CCTGGAAGGG TTTATTGCTA AGTTGGAAGA AATGGGAGTG AGAATGACTG 3660

TATCTGAAGA CAGCATTTTT GTCGAGGAAC AGTCTAATTT GAAAGCAATC AATATTAAGA 3720

CAGCTCCTTA CCCAGGCTTT GCAACTGATT TGCAACAACC GCTTACCCCT CTTTTACTAA 3780

GAGCGAATGG TCGTGGTACA ATTGTCGATA CGATTTACGA AAAACGTGTA AATCATGTTT 3840

TTGAACTAGC AAAGATGGAT GCGGATATTT CGACAACAAA TGGTCATATT TTGTACACGG 3900

GTGGACGTGA TTTACGTGGG GCCAGTGTTA AAGCGACCGA CTTAAGAGCT GGGGCTGCAC 3960

TAGTCATTGC TGGGCTTATG GCTGAAGGTA AAACTGAAAT TACCAATATC GAGTTTATCT 4020

TACGTGGTTA TTCTGATATT ATCGAAAAAT TACGTAATTT AGGAGCGGAT ATTAGACTTG 4080

TTGAGGATTA AACCGTAGAG GTGTTTATGA ATATTTGGAC CAAATTAGCA ATGTTTTCTT 4140

TTTTTGAAAC GGATCGCTTG TATTTGCGTC CTTTCTTTTT TAGTGATAGT CAGGACTTCC 4200

GCGAGATAGC TTCAAATCCA GAAAATCTTC AATTTATTTT CCCAACGCAG GCAAGTCTGG 4260

AAGAAAGTCA ATATGCACTG GCCAATTACT TTATGAAGTC CCCTTTGGGA GTGTGGGCAA 4320

TTTGTGACCA GAAAAATCAA CAAATGATTG GTTCTATTAA ATTTGAGAAG TTAGATGAAA 4380

TCAAAAAAGA AGCTGAGCTT GGCTATTTTT TGAGAAAAGA TGCTTGGTCG CAAGGATTTA 4440

TGACAGAGGT TGTTAGAAAA ATTTGTCAGC TTTCTTTTGA GGAATTTGGC TTAAAACAAT 4500

TATTTATCAT TACCCACCTT GAAAATAAAG CTAGCCAAAG AGTTGCTCTT AAGTCTGGAT 4560

TTAGTTTGTT CCGTCAGTTT AAGGGAAGTG ATCGTTACAC AAGAAAAATG CGGGATTATC 4620

TTGAATTTCG GTATGTAAAA GGAGAGTTCA ATGAGTAAGC ATCAGGAAAT TCTAAGCTAT 4680

TTGGAGGAAT TACCAGTAGG TAAAAGGGTC AGTGTTCGTA GCATTTCGAA TCATCTAGGA 4740

GTTAGTGATG GAACAGCCTA TCGGGCTATT AAAGAAGCTG AAAACCGTGG AATTGTGGAG 4800

ACCCGTCCTA GAAGTGGAAC AATTCGTGTT AAATCCCAGA AAGTTGCTAT AGAGAGATTA 4860

ACGTTTGCTG AAATTGCAGA AGTGACTTCT TCTGAGGTTC TGGCTGGGCA AGAAGGTTTA 4920

GAGAGAGAAT TTAGTAAGTT TTCAATTGGT GCCATGACTG AACAAAATAT CTTGTCTTAC 4980

CTTCATGATG GGGGGCTCTT GATTGTCGGA GACCGAACCC GTATTCAGTT GCTAGCCTTG 5040

GAAAATGAAA ATGCAGTTCT GGTTACAGGG GGATTTCAGG TTCATGATGA TGTGCTTAAA 5100

CTGGCCAATC AAAAAGGGAT TCCTGTTCTA AGAAGTAAGC ATGATACCTT TACCGTCGCG 5160

ACCATGATCA ATAAAGCCTT GTCAAATGTC CAAATCAAGA CTGATATTCT GACAGTTGAG 5220

AAACTTTATC GCCCTAGTCA TGAGTATGGT TTTCTGAGAG AGACAGATAC AGTTAAAGAT 5280

TATTTGGACT TGGTTCGTAA GAATCGTAGC AGCCGTTTCC CTGTTATCAA TCAACATCAG 5340

GTCGTTGTTG GTGTTGTAAC CATGAGAGAC GCTGGTGATA AATCACCAAG CACGACAATT 5400

GATAAGGTTA TGTCTCGTAG TCTATTTTTG GTTGGATTAT CGACAAATAT TGCCAATGTG 5460

AGTCAACGGA TGATCGCAGA AGACTTTGAA ATGGTACCAG TTGTTCGAAG CAATCAAACT 5520

TTGCTTGGCG TTGTGACGCG ACGAGATGTC ATGGAGAAGA TGAGCCGTTC CCAAGTTTCG 5580

GCTCTACCAA CTTTTTCTGA GCAGATTGGA CAAAAGCTCT CTTATCACCA TGATGAAGTA 5640

GTCATTACAG TGGAACCCTT TATGCTAGAA AAAAATGGAG TTTTGGCTAA TGGTGTATTG 5700

GCAGAAATTC TGACCCACAT GACCCGATTT AGTTGTTAAT AGTGGTCGCA ATCTCATTAT 5760

CGAGCAGATG CTGATCTACT TTTTGCAGGC TGTTCAGATA GATGATATAT TGCGCATTCA 5820

GGCACGGATT ATTCATCATA CGAGACGGTC AGCTATAATT GATTACGATA TTTATCATGG 5880

TCACCAGATT GTTTCAAAAG CAAATGTGAC TGTTAAAATT AATTAGAAAC TAGGAGAAAA 5940

GATGATAACA TTAAAATCAG CTCGTGAAAT CGAAGCTATG GACAAGGCTG GTGATTTTCT 6000

AGCAAGTATT CATATAGGCT TACGTGATTT GATTAAGCCA GGCGTAGATA TGTGGGAAGT 6060

TGAAGAATAT GTCCGCCGTC GTTGTAAAGA AGAAAATTTC CTTCCACTTC AGATTGGGGT 6120

TGACGGTGCC ATGATGGACT ATCCTTATGC TACCTGTTGC TCTCTTAACG ATGAAGTGGC 6180

TCACGCTTTC CCTCGTCATT ATATCTTGAA AGATGGTGAT TTGCTCAAAG TTGATATGGT 6240

TTTGGGAGGT CCCATTGCTA AATCTGACCT AAATGTCTCA AAATTAAACT TCAACAATGT 6300

TGAACAAATG AAAAAATACA CTCAGAGCTA TTCTGGTGGT TTAGCAGACT CATGTTGGGC 6360

TTATGCTGTT GGTACACCGT CCGAAGAAGT CAAAAACTTG ATGGATGTAA CCAAAGAAGC 6420

TATGTACAAG GGTATTGAGC AAGCTGTTGT TGGAAATCGT ATCGGTGATA TCGGTGCGGC 6480

TATTCAAGAA TACGCTGAAA GTCGTGGTTA CGGTGTAGTG CGTGATTTGG TTGGTCATGG 6540

TGTTGGCCCA ACTATGCACG AAGAACCAAT GGTTCCTAAC TATGGTATTG CAGGTCGTGG 6600

ACTCCGTCTT CGTGAAGGAA TGGTCTTAAC CATTGAACCA ATGATCAATA CAGGCGATTG 6660

GGAAATTGAT ACAGATATGA AAACTGGTTG GGCGCATAAG ACCATTGACG GTGGATTGTC 6720

ATGTCAGTAT GAACACCAAT TTGTCATTAC GAAAGATGGA CCTGTTATCT TGACTAGCCA 6780

AGGTGAAGAA GGAACTTATT AATAAAAAGT GAAAAGACTA CTGGAAGTTT ATTTTGATAA 6840

AAAATCCAGT AGATCTTTTC ATAATAAAAC GCATTGTATC AAGTGTTAGG GGCTGATATC 6900

ATGCGTTTTT CTGCTTTTAA GATTTTTTCC AACTCTGTTT GTAAGCGCAT CATAACAAAG 6960

GGTCTAGGAT TCAGGGCTCT CCTCCTATAT ACTATTAGTA AAGTAAAACT AAGGGAGGAT 7020

ATTTTAGTGT CGCAGTCTAT TGTTCCTGTA GAGATTCCAC AATATTGTCG TTTTGATTCT 7080

AAAAAGAGAA ATGGAATTCT GTTTAATGTT CGTATTGCCA ATCTTAAATT TACTTTTTTA 7140

TATTATACTT CCTGCGAAAC AAAATATGGT ATAGTAGTTC TATGAATGAT GAAGCAAGTA 7200

AACAACTAAC TGATGCACGA TTTAAGCGTC TTGTTGGTGT TCAGCGTACC ACTTTTGAAG 7260

AGATGTTAGC TGTATTAAAA ACAGCTTATC AACTTAAACA CGCAAAAGGT GGACGAAAAC 7320

CTAAATTAAG CCTAGAAGAC CTTCTTATGC CCACTCTTCA ATAGTGCGAG AATATCGAAC 7380

TTATGAAGAA ATTGCGGCTG ATTTTGGTAT TCACGAAAGC AACTTTATCC GTCGGAGCCA 7440

ATGGGTTGAA ATAACTCTTG TTCAAAGTGG TTTTACGGTT TCAAGAACTC CTCTCAGTTC 7500

TGAGGACACG GTAATGATTG ATGCGACGGA AGTAAAAATC AATCGCCCTA AAAAAACAAT 7560

TAGCGAATGA TTCTGGTAAA AAGAAATTTC ACGCTATGAA GGCTCAAGCG ATTGTCACAA 7620

GTCAAGGGAG AATTGTTTCT TTGGATATCG CTGTGAACTA TAGTCATGAT ATGAAGTTGT 7680

TCAAAATGAG TCGTAGAAAT ATCGAACAAG CTGGTAAAAT CTTGGCTGAC AGTGGTTATC 7740

AAGGGCTCAT GAAGATATAT CCTCAAGCAC AAACTCCACG TAAATCCAGC AAACTCAAGC 7800

CGCTAACAGC TGAAGATAAA GCCTATAACC ATGCGCTATC TAAGGAAAGA AGCAAGGTTG 7860

AGAACATCTT TGCCAAAGTA AAAACGTTTA AAATATTTTC AACAACCTAT CGAAATCATC 7920

GTAAACGCTT CGGATTACGA ATGAATTTGA GTGCTGGTAT TATCAATCAT GAACTAGGAT 7980

TCTAGTTTTG CAGGAAGTCT ATTGAGGTAT TGAGCTAGTT TATGAAAAAA TTGGGTGAAA 8040

AGTCGAGTGT TTTAGAAACC CACAGTGTAG TATTCTAGTT TCAATCCACT ATATTTTGCT 8100

ACTCCCCGTA AAGTTTCTAT TTTCCCTGAT TTCTGATATA ATAGAAATAT TGACTTCAAG 8160

AGTAAGGAAG AGAAGATGAA CGCATTATTA AATGGAATGA ATGACCGTCA GGCTGAGGCG 8220

GTGCAAACGA CAGAAGGTCC CTTGCTAATC ATGGCAGGGG CTGGTTCTGG AAAGACTCGT 8280

GTTTTGACCC ACCGTATCGC TTATTTGATT GATGAAAAGC TGGTCAATCC TTGGAATATC 8340

TTGGCCATTA CCTTTACCAA CAAGGCTGCG CGTGAGATGA AAGAGCGTGC TTATAGCCTC 8400

AATCCAGCGA CTCAGGACTG TCTGATTGCG ACCTTCCACT CCATGTGTGT GCGTATTTTG 8460

CGTCGCGATG CGGACCATAT TGGCTACAAT CGTAATTTTA CAATTGTGGA TCCTGGTGAA 8520

CAGCGAACGC TCATGAAACG TATTCTCAAA CAGTTGAACT TGGACCCTAA AAAATGGAAT 8580

GAACGAACTA TTTTGGGGAC CATTTCCAAT GCTAAGAATG ATTTGATTGA TGATGTTGCT 8640

TATGCTGCCC AAGCTGGCGA TATGTATACG CAAATTGTGG CCCAGTGTTA TACAGCCTAT 8700

CAAAAAGAAC TTCGTCAGTC TGAATCCGTT GACTTTGATG ATTTGATTAT GCTGACCTTG 8760

CGTCTCTTTG ATCAAAATCC TGATGTTTTG ACCTACTACC AGCAAAAATT CCAATACATC 8820

CACGTTGATG AGTACCAAGA TACCAACCAC GCTCAGTACC AATTGGTCAA ACTCTTGGCT 8880

TCCCGTTTTA AAAATATCTG TGTGGTTGGG GATGCGGACC AGTCTATCTA CGGTTGGCGT 8940

GGTGCTGATA TGCAGAATAT CTTGGACTTT GAAAAGGATT ACCCCAAAGC CAAGGTTGTT 9000

TTGTTGGAGG AAAATTACCG CTCAACCAAA ACCATTCTCC AAGCGGCCAA CGAGGTTATT 9060

AAAAATAATA AAAATCGCCG TCCTAAAAAT CTCTGGACTC AAAACGCTGA TGGGGAGCAA 9120

ATCGTTTACT ATCGTGCCGA TGATGAGCTG GATGAGGCTG TATTTGTAGC CAGAACCATC 9180

GATGAACTTA GTCGCAGTCA AAACTTCCTT CATAAGGATT TTGCAGTTCT CTATCGGACT 9240

AATGCCCAGT CCCGTACAAT TGAGGAAGCC CTGCTCAAGT CTAACATTCC TTATACCATG 9300

GTTGGCGGAA CCAAATTCTA CAGCCGTAAG GAAATTCGCG ATATTATTGC TTATCTCAAC 9360

CTTATTGCTA ATTTGAGTGA CAATATTAGT TTTGAGCGTA TTATCAACGA GCCTAAACGT 9420

GGAATTGGTC TAGGTACAGT TGAGAAAATC CGTGATTTTG CAAATTTGCA AAATATGTCT 9480

ATGCTGGATG CTTCTGCTAA TATTATGTTG TCTGGTATCA AGGGTAAGGC AGCCCAATCT 9540

ATCTGGGATT TTGCCAATAT GATGCTTGAT TTGCGGGAGC AGCTAGACCA CTTAAGCATT 9600

ACAGAGTTGG TTGAGTCCGT CCTAGAAAAA ACAGGTTATG TCGATATTCT TAACTCCCAA 9660

GCGACTCTAG AAAGCAAGGC ACGGGTTGAA AATATCGAAG AGTTTCTTTC TGTTACGAAG 9720

AACTTTGATG ACACCACGGA TGTGACAGAA GAGGAAACTG GTCTGGACAA ACTGAGTCGT 9780

TTCTTAAATG ACTTGGCTTT GATTGCCGAC ACAGATTCAG GTAGTCAGGA GACATCAGAA 9840

GTGACCTTGA TGACCCTGCA TGCTGCCAAA GGTCTCGAAT TTCCAGTTGT CTTTTTGATT 9900

GGGATGGAAG AAAATGTCTT TCCACTTAGT CGTGCGACTG AAGATTCAGA TGAATTAGAA 9960

GAAGAGCGCC GTCTAGCCTA TGTAGGTATC ACGCGTGCAG AGAAAATTCT CTATCTGACC 10020

AATGCCAACT CACGCTTGCT TTTTGGTCGT ACCAATTATA ACCGTCCGAC TCGTTTTATT 10080

AACGAAATCA GTTCAGACTT GCTTGAGTAT CAAGGTCTGG CTCGTCCTGC AAATACAAGC 10140

TTTAAGGCAT CATATAGCAG TGGTAGTATT TCCTTTGGTC AAGGTATGAG TTTGGCTCAG 10200

GCTCTTCAAG ACCGTAAACG CGGTGCTGCC CCAAAATCAA TCCAGTCAAG CGGTCTTCCA 10260

TTTGGTCAAT TTACAGCTGG CGCAAAACCA GCATCTAGCG AGGCAAATTG GTCCATTGGT 10320

GATATTGCTC TCCACAAGAA ATGGGGAGAG GGAACCGTTC TGGAAGTTTC AGGTAGCGGT 10380

GCTAGGCAGG AATTGAAAAT CAATTTCCCA GAAGTAGGTT TGAAAAAACT TTTAGCCAGT 10440

GTGGCTCCAA TTGAGAAAAA AATCTAATTT TCCATCCTTC TCACGAATAA TAAAGTGAGG 10500

AGGATTTTTA TGTACAGTAT TTCATTCCAA GAAGATTCAC TATTACCAAG AGAAAGGCTG 10560

GCCAAGGAAG GAGTTGAAGC GCTTAGTAAC CAAGAGTTGC TAGCTATTTT ACTCAGGACA 10620

GGAACACGTC AAGCTAGCGT TTTTGAAATT GCCCAAAAAG TCTTGAACAA TCTTTCAAGC 10680

CTAACGGATT TGAAAAAAAT GACCCTGCAG GAATTGCAGA GTTTGTCTGG TATTGGGCGT 10740

GTTAAGGCCA TAGAATTACA AGCTATGATT GAACTGGGGC ATCGTATTCA CAAACACGAG 10800

ACTCTTGAAA TGGAAAGTAT TCTCAGCAGT CAAAAGTTGG CCAAGAAGAT GCAGCAGGAA 10860

TTAGGGGATA AAAAACAAGA GCACCTGGTG GCACTCTATC TCAATACTCA AAATCAAATC 10920

ATCCATCAGC AGACCATTTT TATCGGGTCT GTAACTCGTA GTATCGCTGA ACCGCGAGAG 10980

ATTCTTCACT ATGCAATCAA GCATATGGCG ACTTCTCTTA TCTTGGTCCA CAATCATCCT 11040

TCAGGAGCGG TAGCGCCTAG CCAAAATGAT GATCATGTCA CTAAACTTGT TAAAGAAGCC 11100

TGCGAATTGA TGGGGATTGT TCTCTTGGAC CATTTGATTG TCTCTCATTC TAATTACTTT 11160

AGTTATCGTG AAAAGACAGA TTTAATCTAA AGTTCATTAA CGACATAGTC AAAGAGTTTT 11220

TTATCTTTGG GACGATTTTC AAAAAGAAGT TCTGGATGCC ATTGGACACC GAGAAAGGCG 11280

ACATCATCCG TACTCATGAC AGCCTCAATG ATACCATCTT TAGGATCATG AGCCACAACT 11340

TTTAAATTTG GTGCTAAGTC CTTGATGCTC TGGTGGTGGA AGGAGTTGAT ATGAGAGATT 11400

TCTCCATAGA TTTCTTGGAG AACGGTATCT GGTTCTGTTA CCAAGCGTTG AGTTGTGTAC 11460

TCAACAGAAG AATCCTGCCA ATGGTCTTCG ATATCTTGGT ACAAAGTTCC ACCCATGGCA 11520

ACGTTAAAGA GTTGGGTACC ACGGCAGACA GAGAAAATGG GCTTTTTCTG TTTAATAGCT 11580

TCCTTGATGA GGGCCAGTTC GAAGATATCT CTTTGAAGGT GATAGTCATC ACTATCAATG 11640

GTTTTGGGTT CGCCATAAAA TTTTGGATCG ACATTTTGCC CACCTGTCAA GATGAGCTTG 11700

TCAATCAAAC TGATATAGTG GCAGGCCATT TCTTGATCAC CAATCGGTAG GATGATGGGA 11760

ATCCCTCCAG CATCTTTAAC GCCTTCAACA AAGCCTTTTG CTGCGTAGCT CATCATGATG 11820

TCATCATCTG GATGAGTTTT TTCGTTTCCT GTAATCCCAA TAACTGGTTT TTTCATAAAA 11880

TGATTTTCGC TTTCTAATCC TCTTTTCGCA TGAAGTAGAG GAGGGTTTGG AGTTCACTTG 11940

TCAAATCGAC ATACTGAACG ACCACGTCTT TTGGTAAATG CAGATGGACT GGTGAAAAAC 12000

TGAGAATTCC TTTCACACCA GCATCAACCA AGAGATTAGC AACCTCTTGT GACTTGACGC 12060

TGGGAACAGT TAGGATAGCA GTCTTCACAT CAGCATCCTT GATTTTATCC TTGATCTGAG 12120

AAATCCCGTA AATGGGAATC CCGTCAGGAG TTTGGGTACC GACTTCAGGA TGGTCGTCTA 12180

GGTCAAAGGC CATGATAATC TTCATCTTGT TACGTTCGTG GAAGCGGTAG TGGAGAAGGG 12240

CATGGCCCAT ATTTCCAATA CCAACCAGCA TGACATTGGT AATAGAGTTG TCATTGAGCA 12300

AATCGGCAAA AAATGTCATT AGTTTTTTGA CATCATAGCC AAAACCACGA CGACCAAGTT 12360

CACCAAAATA GGAAAAATCA CGACGTACGG TCGCTGAATC AATACCGATA GCCTCTGCAA 12420

TTTGCTTAGA GTTGGCACGT TCAATCTTTT CTGCATGAAA TCTCTTAAAA ATTCGATAGT 12480

AGAGAGAGAG TCTTTTTGCT GTAGCTTTTG GAATAGCAAA CTGTTTATCT TTCACAAAAT 12540

CACAACCTTT CTATTCTTCT ATTTTATAGA AACATTGTGA AAAAATCAAC AAAAATAAGA 12600

AAAAACTAAG AAAAATCTTA GTTTTGATGT AAAAAATCTG CATGAGATAG AAAACGGTAG 12660

AGGTCTCCGA CCAGCCCCTG ATAAACTTTT TTGCCCCTAA AAGTCAGAGA AGTCACATAA 12720

AGTGTATCTG GTAAGGTTAC ACATCCTGAC AAAGTCAACA TGAGAGCCTC ATGATCCTCA 12780

TACTTGAGAG TACGCTCTAC ATGATAGCAG TCCTTATAGG TCAGTTCAAA CATTTTGGCT 12840

CTATCTTTCC GATTTTGTAA AGACACCACG TTCTACCAAG CTATCCATGA GGAAGTAGAA 12900

TTTTTCCTGA TGAATATGGT GGTCTTCTGA TTTGAAAATA TCAACTAGAC GAAGGCCAAA 12960

CTTGTCAGTG ATATTGATTT TAGCCCCTGT AAGTTCCTTG TTAATGATGA TTTTGAGTTG 13020

GAAGCCTTCA CCGCTGTTTG GCACTTTTTC CAAAAGGCGA GTCAGTTCAT AGTTACCAAC 13080

CTTAGTTTCA AAAAAGGTGT TATCTTTGAG GGTGAATTTT TTAACAGAAG GGCTAAGAGT 13140

GTAATCGTAA CGACAATTTT TTAACTGAAT GATTTTTTCA AATGCCATAT GGCTAACCTC 13200

CGATAATTTC TTTTAAGGTT TTTGCGAGGG TTTGTAGGTC TTCAACGGTA TTTTGTGGCG 13260

ACAAACTGAT GCGAAGGGAT TCCTTCAAGC GTTCTGAATT TGCGCCATAC ATGGCTTCAA 13320

GAACATGGCT GGATTGGACA ACGCCTGCAG TACAGGCTGA GCCAGTAGAG ATTGAAATTC 13380

CAGCTAAATC TAGCCGAAGG AGTAAGAGGT CATTTTTCTG ACCAGGAAAT CCAATATTGA 13440

GAACATAAGG GAGATGATGT TTTCCTCTAT TCAGGTAATA CTGAATGCCC TCCAGCTCTG 13500

CCAGAAAGGC AGTTTCTAGA TTTTGTACAT GTTGAAAATG TTCTTCTTGT TTTTCTAGGT 13560

CTTCTTTTAG GGCTGCAACC ATGCCTACAA TGGCAGGCAG ATTTTCAGTT CCTGCACGTT 13620

TTTTCTGTTC CTGGTCTCCG CCATGTAGAT AGGAATCAAA GTCCATGCTA GATGCGTAGA 13680

GAAAACCGAT TCCCTTAGGA CCATGGAATT TGTGGGCAGA AGCAGTGAGA AAATCAATGC 13740

CCAATTCTTC TGAATGAATT GGGATTTTAC CAATAGCCTG AACTGCATCA ACATGATAGG 13800

CAGCAGGGTG TTGCTTGAGT ATTTGGCCAA TTTCAGCGAT GGGCAGTAGG TTTCCTGTCT 13860

CATTATTGAC AAACATGGTA GAAACCAAAA TCGTATCGTC ACGTAAAGCC TTTTGAATTT 13920

GCTGGGCTGT GATTTCTTGA TTTTCTGGCT GGATAATGGT TGCTTCAAAC CCAAAGTGTT 13980

GAACCAAGTA ATCAATTGTT TCAAGGACAG CATGGTGCTC GATGGCAGTT GTGATGATAT 14040

GTTTTCCTTG TTCTTGGTGA CGAAGACAGT AGCCAATGAT GGTAGTATTA TTGCCTTCAG 14100

TCCCACCAGA AGTGAAAAAG ATATGTTGAG GTTTTGTCCT TAGTAACTGG GCTAGTTCCT 14160

GACGGGCTTC TCGCAAGAGT TTGCCAGCTT GACGACCATG ACCATGAATA CTAGAAGGAT 14220

TTCCGTGGGT TTCTTGCATA ACCTTGGTCA TAGCTGAAAT AGCAACTGCT GACATAGGAG 14280

TCGTTGCAGC ATTGTCCAAA TAAATCAAAG AATCACCTTA TTTCTTTTTA TTGTAGGCAA 14340

AGAGTGGGCT GACTGGTTTT CTTTCGTGAA TACGGACGAT AGCATCACCA ATTAACTCAC 14400

TAGCAGTGAT GTAGCATACA TTTTTAGGAG TTTTTTCTTT TGTTGCTACT GAATCAGTCA 14460

CAAGAATTTC TTTAATATTA GTATTGTCAA GAAGCTCAGC AGCTCCCTCG ACGAAGAGAC 14520

CGTGGCTAGA AACAGCATAA ATTTCTGTAG CTCCTTCACG TTCAACGATT TTAGAAGCTT 14580

CAGAGAAGGT ACGTCCTGTA TTTAAAATAT CATCAATCAA GATAGCTTTC TTACCTTCAA 14640

CATCACCAAT AATATAACCT TCGTTACGAG TTGCATCGTC TTGAGGGTAG TCGATAATGG 14700

CGATAGGAGC ATCAAGATAT TCAGCCAGGC TACGCGCACG TTTGACACCT GAATTTTTAG 14760

GGCTAACGAC AACAACATCT GAACCAAGCA ATCCTTTATC GCAGTAATGT TTTGCGAATA 14820

GGGGAACAGT GAAAAGATTA TCCACTGGAA TATCAAAGAA ACCTTGAACC TGAACGGCAT 14880

GCAAATCAAG AGTCAGGATA CGATCAACTC CAGCCTTAAC CAGCATATTG GCAACTAGTT 14940

TTGCTGTAAG TGGCTCACGA GGACAAGCAA TGCGGTCTTG ACGTGCATAG CCAAAATATG 15000

GAAGGACAAC GTTGATACTG TGGGCACTTG CACGCACACA AGCATCGACC ATGATTAACA 15060

ATTCCATTAG GTGGTTGTTG ACAGGGAAAC TTGTTGATTG GATGATGTAA ACATCATAAC 15120

CACGGACACT TTCTTCGATA TTTACTTGGA TTTCTCCGTC TGAAAATTGA CGTGATGATA 15180

GTTTTCCAAG TGGGACACCA ACAGCTTGGG CAATTTTTTG TGCAATCTCT TGGTTAGAGT 15240

TGAGTGCGAA AAGTTTCATG TTTTTTCTAT CTGACATTAT AGACCGTCCT CTGTAAACTT 15300

TATAAATCCT AGTTATATTT ACCTTACATA TATGAACTGG GATTTGTGTA TTTTTATCTT 15360

TTCTATTTTA CCAAAAAATG GAGATTATTT CAGCTATTTT TCATACTTTT GACAAATCGA 15420

ACCAATTTTG AAGGAGCTTT TTGATAGGAA ATCTGATTTT TCTCTAAAAA TTGTCGAAAA 15480

TCCTGTTTGC CTTGCTCATG ATTTTCCACT TCAAGCTCCA ATTCGTAATC TGTTATATCA 15540

AAGTATCGGC TCTGATCCAG TGCCATGAGA CCAATAGCTG TTTTCATTTC ATAGCGAAGC 15600

GTTGTTAGAC AACCAAGAAC CTGCCAGTTC TTACTTTGGA TACCATGTTT CGCCAATTCA 15660

TCCAGTACTA GCCCTTGAGG AAGTTCTTCC TTACTCAGAT AGTTCTCAGC ATCTTTTAGT 15720

TGCAATTTTT GGTTGTATTC CATGTTTCCA ACACTCTGCG GGACTTTGAG TGTCAACTCA 15780

GCCCAGTCTT CAAAGGTTCG AATGCGCATA GCGACTTTCT TTTCCTCCAG TTCAAAATCA 15840

GGCGTGTCGA TGTAGTAATT TGTTTGAAGA ACAGGAGTGA CACCTGTGAA CTGGTCTTTT 15900

AGACGATTGT ATTCATCTTT TTTCAATAGT GTTTTCAATT CAATTTCTAA ATGTTTCATT 15960

TTTCTTACCT TTTTTTATCG TTGAAAGCGG ATTTATGGTA TAATAAGCAT TGTATTTATT 16020

GTATATGAAT CTGGAGAAAA AATCAAAGAT ATTTTTGACG GATAATATGA GAACAAGGGA 16080

GAATATATGA CCTTAGAATG GGAAGAATTT CTAGATCCTT ACATTCAAGC TGTTGGTGAG 16140

TTAAAGATTA AACTTCGTGG TATTCGTAAG CAATATCGTA AGCAAAATAA GCATTCTCCA 16200

ATTGAGTTTG TGACCGGTCG AGTCAAGCCA ATTGAGAGCA TCAAAGAAAA AATGGCTCGT 16260

CGTGGCATTA CTTATGCGAC CTTGGAACAC GATTTGCAGG ATATTGCTGG CTTACGTGTG 16320

ATGGTTCAGT TTGTAGATGA CGTCAAGGAA GTAGTGGATA TTTTGCACAA GCGTCAGGAT 16380

ATGCGAATCA TACAGGAGCG AGATTACATT ACTCATAGAA AAGCATCAGG CTATCGTTCC 16440

TATCATGTGG TAGTAGAATA TACGGTTGAT ACCATCAATG GAGCTAAGAC TATTTTGGCA 16500

GAAATTCAAA TTCGTACTTT GGCCATGAAT TTCTGGGCAA CGATAGAACA TTCTCTCAAC 16560

TACAAGTACC AAGGGGATTT CCCAGATGAG ATTAAGAAGC GACTGGAAAT TACAGCTAGA 16620

ATCGCCCATC AGTTGGATGA AGAAATGGGT GAAATTCGTG ATGATATCCA AGAAGCCCAG 16680

GCACTTTTTG ATCCTTTGAG TAGAAAATTA AATGACGGTG TAGGAAACAG TGACGATACA 16740

GATGAAGAAT ACAGGTAAAC GAATTGATCT GATAGCCAAT AGAAAACCGC AGAGTCAAAG 16800

GGTTTTGTAT GAATTGCGAG ATCGTTTGAA GAGAAATCAG TTTATACTCA ATGATACCAA 16860

TCCGGATATT GTCATTTCCA TTGGCGGGGA TGGTATGCTC TTGTCGGCCT TTCATAAGTA 16920

CGAAAATCAG CTTGACAAGG TCCGCTTTAT CGGTCTTCAT ACTGGACATT TGGGCTTCTA 16980

TACAGATTAT CGTGATTTTG AGTTGGACAA GCTAGTGACT AATTTGCAGC TAGATACTGG 17040

GGCAAGGGTT TCTTACCCTG TTCTGAATGT GAAGGTCTTT CTTGAAAATG GTGAAGTTAA 17100

GATTTTCAGA GCACTCAACG AAGCCAGCAT CCGCAGGTCT GATCGAACCA TGGTGGCAGA 17160

TATTGTAATA AATGGTGTTC CCTTTGAACG TTTTCGTGGA GACGGGCTAA CAGTTTCGAC 17220

ACCGACTGGT AGTACTGCCT ATAACAAGTC TCTTGGCGGT GCTGTTTTAC ACCCTACCAT 17280

TGAAGCTTTG CAATTAACGG AAATTGCCAG CCTTAATAAT CGTGTCTATC GAACACTGGG 17340

CTCTTCCATT ATTGTGCCTA AGAAGGATAA GATTGAACTT ATTCCAACAA GAAACGATTA 17400

TCATACTATT TCGGTTGACA ATAGCGTTTA TTCTTTCCGT AATATTGAGC GTATTGAGTA 17460

TCAAATCGAC CATCATAAGA TTCACTTTGT CGCGACTCCT AGCCATACCA GTTTCTGGAA 17520

CCGTGTTAAG GACGCCTTTA TCGGCGAGGT GGATGAATGA GGTTTGAATT TATCGCAGAT 17580

GAACATGTCA AGGTTAAGAC CTTCTTAAAA AAGCACGAGG TTTCTAAGGG ATTGCTGGCC 17640

AAGATTAAGT TTCGAGGTGG AGCTATTCTG GTCAATAATC AACCGCAAAA TGCAACGTAT 17700

CTATTGGACG TTGGAGACTA CGTTACCATT GACATTCCCG CTGAGAAAGG CTTTGAAACC 17760

TTGGAGGCTA TTGAGCTTCC ATTAGATATT CTCTATGAGG ATGACCACTT TCTAGTCTTG 17820

AATAAACCCT ATGGAGTGGC TTCTATTCCT AGTGTCAATC ACTCTAATAC CATTGCCAAT 17880

TTTATCAAGG GTTACTATGT CAAGCAAAAT TATGAAAATC AGCAGGTTCA CATTGTTACC 17940

AGACTAGATA GGGATACTTC TGGCTTGATG CTCTTTGCCA AGCACGGTTA TGCCCATGCA 18000

CGATTAGACA AGCAGTTGCA GAAGAAATCT ATCGAGAAAC GCTACTTTGC TTTGGTTAAG 18060

GGAGATGGAC ATTTGGAGCC AGAAGGGGAA ATTATTGCTC CGATTGCGCG TGATGAAGAT 18120

TCCATTATTA CCAGACGAGT GGCTAAAGGC GGAAAGTATG CCCATACTTC ATACAAGATT 18180

GTAGCTTCTT ATGGAAATAT TCACTTGGTC TATATTCACC TGCACACTGG TCGAACCCAT 18240

CAAATCCGAG TCCATTTTTC TCATATCGGT TTTCCTTTGC TGGGAGATGA TTTGTATGGT 18300

GGTAGTCTGG AAGATGGTAT TCAACGTCAG GCTCTGCATT GCCATTACCT ATCCTTTTAT 18360

CATCCATTTT TAGAGCAAGA CTTGCAGTTA GAAAGTCCCT TGCCGGATGA TTTTAGTAAC 18420

CTTATTACCC AGTTATCAAC TAATACTCTA TAAAAACTGT CTCAGAGTAT AATTATTATC 18480

TTAAAGGAGA AAACTCATGG AAGTTTTTGA AAGTCTCAAA GCCAACCTTG TTGGTAAAAA 18540

TGCTCGTATC GTTCTCCCTG AAGGGGAAGA GCCTCGTATT CTTCAAGCAA CAAAACGCTT 18600

AGTAAAAGAA ACAGAAGTGA TTCCTGTTTT GCTTGGAAAT CCTGAAAAAA TTAAAATTTA 18660

TCTTGAAATT GAAGGAATCA TGGATGGTTA TGAGGTCATC GACCCTCAAC ATTATCCTCA 18720

ATTTGAAGAA ATGGTTTCTG CCTTGGTGGA GCGTCGCAAG GGCAAAATGA CTGAAGAAGA 18780

TGTACGCAAG GTTTTGGTTG AAGATGTCAA CTACTTTGGT GTGATGTTGG TTTACTTGGG 18840

CTTGGTTGAT GGAATGGTGT CAGGAGCGAT TCACTCAACA GCTTCAACAG TTCGCCCAGC 18900

TCTACAAATC ATCAAAACTC GTCCAAATGT AACTCGTACT TCAGGAGCCT TCCTCATGGT 18960

TCGTGGTACG GAACGTTACC TATTTGGAGA CTGTGCCATT AACATCAATC CAGATGCAGA 19020

AGCCTTGGCT GAAATTGCCA TCAACTCAGC AATCACAGCT AAGATGTTTG GCATCGAACC 19080

TAAAATTGCC ATGTTGAGCT ATTCTACTAA AGGTTCAGGG TTTGGTGAAA GCGTTGATAA 19140

GGTCGTTGAA GCAACTAAAA TTGCTCACGA CTTGCGTCCT GACCTTGAAA TCGATGGTGA 19200

GTTGCAATTT GATGCAGCCT TTGTTCCTGA AACTGCAGCT CTGAAAGCTC CTGGAAGTAC 19260

GGTAGCTGGT CAAGCAAATG TCTTCATCTT CCCAGGTATC GAGGCAGGAA ATATTGGTTA 19320

CAAGATGGCT GAACGCCTGG GTGGCTTTGC GGCTGTAGGA CCTGTTTTGC AAGGTTTAAA 19380

CAAGCCAGTT AATGATCTTT CTCGTGGATG TAATGCAGAT GATGTTTACA AGTTGACCCT 19440

CATCACAGCA GCTCAAGCAG TTCATCAATA GTGAAAACTA TAAAGTGATA TACTATGCTA 19500

TACTGTAGTT ATGAAACTAT GTACGAAAAG CACTGCCATT AATTCCTGAG AACTAAATTA 19560

CTGATTGGTG TCAAAAAGGA AAACTTCCAA GCGATGATAT CCTGTCTATA CACGACCTAT 19620

AGAAATCTGT AATATACATA TCCGTAAAAC GATAAATTCC CTTTTTGATT TTAAATGAGT 19680

ATGAAAAGAG AATTTTTTGG CTCTTTGTCA ACTGTAGTGG GTTGAAGAAA AGCTAAGCTC 19740

GAGAAAGGAC AAATTTCATC CTTTCTTTTT TGATATTCAG AGCGATAAAA ATCCGTTTTT 19800

TGAAGTTTTC AAAGTTCCGA AAACCAAAGG CATTGCGCTT GATAAGTTTG ATGAGATTAT 19860

TGGTCGCTTC CAGTTTGGCG TTAGAATAGT GTAGTTGAAG GGCGTTGATA ATCTTTTCTT 19920

TATCTTTGAG GAAGGTTTTA AAGACAGTCT GAAAAATAGG ATGAACCTGC TTAAGATTGT 19980

CCTCAATAAG TCCGAAAAAT TTCTCTGGTT CCTTATTCTG GAAGTGAAAA AGCAAGAGTT 20040

GATAGAGCTG ATAGTGGTGT TTCAAGTCTT CCGAATAGCT CAAAAGCTTG TTTAAAATCT 20100

CTTTATTGGT TAAGTGCATA CGAAAAATAG GACGATAAAA TCGCTTATCA CTCAGTTTAC 20160

GGCTATCCTG TTGAATGAGT TTCCAGTAGC GCTTGATAG 20199

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19702 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ACCCGATGTA TCAGCGGATA TTTACTCTAT TTTTCAAACG ATGTTATACC CACAATAAAA 60

GAAAAAAGAC CCTAAGGTCT CCTTTGCTTT TATTATTAAA CGCGTTCAAC TTTACCTGAT 120

TTCAAAGCAC GAGCTGAAGC CCAAACTTTT TTAGGTTTAC CATCGATAAG AACAGTAACT 180

TTTTGAAGGT TTGGTTTTAC GGCACGTTTT GTTTGGTTCA TCGCGTGTGA ACGGTTGTTT 240

CCTGATACAG TCTTACGACC TGTAAAGTAA CATACTTTAG CCATTGTGTT TTCCTCCTAT 300

TAGATCTAAT ATAGCGGATG TGCTAGCACC ACATACCGTA CTATGTTATC ACATTTTCTT 360

GTTTTTTGCA AGGGAATTGG AAGATTTTTT ATTTGTGTCT TAAATCAGGT CTTGCGTGAC 420

ATTTCTGCTC TCCACATGCC ATCGTTGATT AACAGAACAC CAGAATTAAA ATTATGTGTA 480

TAAAAATCAT CTCTAACTGC AGCTAAGGGT ATAGCCGTCA AGTCCAAATC CCACAGCTCA 540

TCTATCGATT TTCTTACAAC AATATCTGAA TCCAAATACA GTACACGAGA CTCGCTTACA 600

TACTTTGGAA TAAAATACCT AAAAAAGCCG CATATGAAAG TCCCTCAAAG GGGAGACGAT 660

AACCTTTCAG AATATTACTG TCAATCTAAA CATTCACAAT CTCACTATTC AAAGTCTCTA 720

GTCTTTTTTC CATCAATTGG AACCATTCTC GCGGAAGGTC ATCATTAAAA ACATAAAACT 780

TAAGATTATA ATGATGAACA CAAAGAGATT TTATTGTTGT TTCAACTTTA TCCATATAAG 840

CATTATCTGC ACCTAAGACA ATCGCTTTTT TCTCTTCTTT CACTTTTTAT CTCATTTCTT 900

TTTATTCCCA TCATATTATT CCCATCATAT GTTTCCCATC ATATGTTTCT ACGTAACCAT 960

TATTTTCGCC TATTCGTTCG TAAAACCATA CCAGTGGAGA TTTTAGATGA AGTCCCATTA 1020

CGGTTTACAA TTTTTACATT ACGACACGGA GTTTTACAAA TCGATTTCAT TTGCCAAACG 1080

TAGTTAGTGA GGCAGTTAGC TAGTTCGCCA AATAGCGACT AGCGTCCAAC AATTTGGAAC 1140

TTTAGTTCCA ATTGTTGGTA CTGAGTCACA TCTTCTCCTC TAACTCTACG TCTGGATACT 1200

TGTCCGCAAA CCAGCGGAGG GCAAAGTCAT TTTCAAAGAG AAAGACTGGT TGGTCAAAAC 1260

GGTCTTTGGC TAAGATATTG CGACTTGACG ACATCCGTTC ATCCAAGTCC TCAGGCTTGA 1320

TCCAACGAAC GGTCTTTTTA CCCATTGGGT TCATAACTAC TTCCGCATTG TACTCGCCTT 1380

CCATGCGGTG TTTAAAGACT TCAAACTGGA GTTGACCTAC AGCGCCTAGC ATGTACTCAC 1440

CTGTTTGGTA ATTCTTATAA AGCTGAACGG CTCCTTCTTG CACCAATTGC TCAATCCCCT 1500

TGTGGAAGGA TTTTTGCTTC ATAACATTCT TAGCAGAAAC TTTCATGAAA ATCTCAGGTG 1560

TAAAGGTTGG CAGGGGTTCA AATTCAAACT TGTTTTTTCC AACCGTCAAG GTATCCCCAA 1620

CCTGATAAGT ACCGGTATCG TAAACCCCGA TAATATCACC TGCCACGGCA TTGGTCACAT 1680

TCTCACGACT CTCCGCCATA AACTGGGTAA CATTAGATAG TTTAGGCGCG TTACGAGTAC 1740

GAGGGAGATT GACACTCATG CCGCGCTCAA ATTCGCCAGA TACGATACGG ACAAAGGCAA 1800

TACGGTCACG GTGACGAGGG TCCATGTTGG CTTGGATTTT AAAGACAAAG CCTGAGAAAT 1860

CCTTGTCATA AGGATCCACA ATTTCACCGT CTGTTTTCTT GTGACCATGT GGTTCTGGAG 1920

CAAACTTGAG GAAGGTTTCA AGGAAGGTCT GCACACCAAA GTTTGTCAGG GCTGAACCGA 1980

AAAAGACAGG CGTCAATTCT CCAGCCAGAA TAGCTTCCTC TGAAAACTCA TTCCCGGCTT 2040

CATTTAAAAG CTCAATGTCA TCCTTGACTT GCTCGTAGAA AGGATTGCTA CCAAAGAGTT 2100

TGTCCCCGTC TTCTAGACTG GCAAAACGCT CATCCCCTTT GTAAAGCTCT AAACGTTGGT 2160

TATAGAGGTC ATACAAGCCC TCAAAGGCTT TCCCCATCCC GATAGGCCAG TTCATAGGGT 2220

AGCTAGCAAT GCCCAAGATT TCTTCCAATT CTTGCAAGAG ATCCAAAGGC TCACGACCGT 2280

CACGGTCCAG CTTGTTCATA AAGGTAAAGA CTGGAATGCC ACGATGTTTC ACAACCTCAA 2340

ACAATTTCTT GGTTTGAGCC TCGATCCCCT TGGCAGAGTC CACGACCATG ACCGCAGCAT 2400

CCACCGCCAT CAAGGTACGA TAGGTATCTT CTGAGAAGTC CTCGTGCCCT GGCGTGTCTA 2460

AGATATTCAC GCGCTTGCCG TCGTAGTCAA ATTGCATAAC AGATGAAGTA ACAGAAATCC 2520

CACGTTGCTT CTCGATATCC ATCCAGTCAG ATTTAGCAAA AGTCCCTGTT TTCTTCCCTT 2580

TTACCGTACC AGCCTCACGA ATCTCACCCC CAAAGTAGAG TAACTGCTCA GTGATGGTTG 2640

TTTTCCCCGC GTCCGGGTGG GAGATAATGG CAAAGGTACG ACGTTTCTTA ATTTCTTCTT 2700

GAATATTCAT AAGTTCTCTT TCTTTGATTC TCTATTTTTC TTGTTTCAAT AGCTGAGAAT 2760

GATTTTTACA TTGGATTTTA CCATTCCTTT CAACACTCCA TTATATCGGA TTTTAGCATT 2820

TTTTTCAATT TCTATTTCTT TTCACTTCCC CCTCCCTTAT TTATAGGAAA ATATGGTAAA 2880

ATAGAACAGA CTAAAAATCA TCATTTCACG AAAGGATGCA AGATGAAAAT TACGCAAGAA 2940

GAGGTAACAC ACGTTGCCAA TCTTTCAAAA TTAAGATTCT CTGAAGAAGA AACTGCTGCC 3000

TTTGCGACCA CCTTGTCTAA GATTGTTGAC ATGGTTGAAT TGCTGGGCGA AGTTGACACA 3060

ACTGGTGTCG CACCTACTAC GACTATGGCT GACCGCAAGA CTGTACTCCG CCCTGATGTG 3120

GCCGAAGAAG GAATAGACCG TGATCGCTTG TTTAAAAACG TACCTGAAAA AGACAACTAC 3180

TATATCAAGG TGCCAGCTAT CCTAGACAAT GGAGGAGATG CCTAATGACT TTTAACAATA 3240

AAACTATTGA AGAGTTGCAC AATCTCCTTG TCTCTAAGGA AATTTCTGCA ACAGAATTGA 3300

CCCAAGCAAC ACTTGAAAAT ATCAAGTCTC GTGAGGAAGC CCTCAATTCA TTTGTCACCA 3360

TCGCTGAGGA GCAAGCTCTT GTTCAAGCTA AAGCCATTGA TGAAGCTGGA ATTGATGCTG 3420

ACAATGTCCT TTCAGGAATT CCACTTGCTG TTAAGGATAA CATCTCTACA GACGGTATTC 3480

TCACAACTGC TGCCTCAAAA ATGCTCTACA ACTATGAGCC AATCTTTGAT GCGACAGCTG 3540

TTGCCAATGC AAAAACCAAG GGCATGATTG TCGTTGGAAA GACCAACATG GACGAATTTG 3600

CTATGGGTGG TTCAGGTGAA ACTTCACACT ACGGAGCAAC TAAAAACGCT TGGAACCACA 3660

GCAAGGTTCC TGGTGGGTCA TCAAGTGGTT CTGCCGCAGC TGTAGCCTCA GGACAAGTTC 3720

GCTTGTCACT TGGTTCTGAT ACTGGTGGTT CCATCCGCCA ACCTGCTGCC TTCAACGGAA 3780

TCGTTGGTCT CAAACCAACC TACGGAACAG TTTCACGTTT CGGTCTCATT GCCTTTGGTA 3840

GCTCATTAGA CCAGATTGGA CCTTTTGCTC CTACTGTTAA GGAAAATGCC CTCTTGCTCA 3900

ACGCTATTGC CAGCGAAGAT GCTAAAGACT CTACTTCTGC TCCTGTCCGC ATCGCCGACT 3960

TTACTTCAAA AATCGGCCAA GACATCAAGG GTATGAAAAT CGCTTTGCCT AAGGAATACC 4020

TAGGCGAAGG AATTGATCCA GAGGTTAAGG AAACAATCTT AAACGCGGCC AAACACTTTG 4080

AAAAATTGGG TGCTATCGTC GAAGAAGTCA GCCTTCCTCA CTCTAAATAC GGTGTTGCCG 4140

TTTATTACAT CATCGCTTCA TCAGAAGCTT CATCAAACTT GCAACGCTTC GACGGTATCC 4200

GTTACGGCTA TCGCGCAGAA GATGCAACCA ACCTTGATGA AATCTATGTA AACAGCCGAA 4260

GCCAAGGTTT TGGTGAAGAG GTAAAACGTC GTATCATGCT GGGTACTTTC AGTCTTTCAT 4320

CAGGTTACTA TGATGCCTAC TACAAAAAGG CTGGTCAAGT CCGTACCCTC ATCATTCAAG 4380

ATTTCGAAAA AGTCTTCGCG GATTACGATT TGATTTTGGG TCCAACTGCT CCAAGTGTTG 4440

CCTATGACTT GGATTCTCTC AACCATGACC CAGTTGCCAT GTACTTAGCC GACCTATTGA 4500

CCATACCTGT AAACTTGGCA GGACTGCCTG GAATTTCGAT TCCTGCTGGA TTCTCTCAAG 4560

GTCTACCTGT CGGACTCCAA TTGATTGGTC CCAAGTACTC TGAGGAAACC ATTTACCAAG 4620

CTGCTGCTGC TTTTGAAGCA ACAACAGACT ACCACAAACA ACAACCCGTG ATTTTTGGAG 4680

GTGACAACTA ATGAACTTTG AAACAGTCAT CGGACTTGAA GTCCACGTAG AGCTCAACAC 4740

CAATTCAAAA ATCTTCTCAC CTACTTCTGC CCACTTTGGA AATGACCAAA ATGCCAACAC 4800

TAACGTGATT GACTGGTCTT TCCCAGGAGT TCTACCAGTT CTCAATAAAG GGGTTGTTGA 4860

TGCCGGTATC AAGGCTGCTC TTGCCCTCAA CATGGACATC CACAAAAAGA TGCACTTTGA 4920

CCGCAAGAAC TACTTCTATC CTGATAACCC CAAAGCCTAC CAAATTTCTC AGTTTGATGA 4980

ACCAATCGGA TATAATGGCT GGATTGAAGT CAAACTAGAA GACGGTACGA CCAAGAAAAT 5040

CGGTATCGAA CGTGCCCACC TAGAGGAAGA CGCTGGTAAA AACACCCATG GTACAGATGG 5100

CTACTCTTAT GTTGACGTCA ACCGCCAAGG GGTTCCCTTG ATTGAGATTG TATCTGAGGC 5160

AGATATGCGT TCTCCTGAAG AAGCCTATGC TTATCTGACA GCCCTCAAGG AAGTTATCCA 5220

GTACGCTGGC ATTTCTGACG TTAAGATGGA GGAAGGTTCG ATGCGTGTGG ATGCCAACAT 5280

CTCCCTTCGT CCTTATGGTC AAGAGAAATT CGGTACCAAG ACTGAATTGA AGAACCTCAA 5340

CTCCTTCTCA AACGTTCGTA AAGGTCTTGA ATACGAAGTC CAACGCCAGG CTGAAATTCT 5400

TCGCTCAGGT GGTCAAATCC GCCAAGAAAC ACGCCGTTAC GATGAAGCGA ATAAAGCAAC 5460

CATCCTCATG CGTGTCAAGG AAGGGGCTGC TGACTACCGC TACTTCCCAG AACCAGACCT 5520

ACCCCTCTTT GAAATTTCTG ACGAGTGGAT TGAGGAAATG CGGACTGAGT TGCCAGAGTT 5580

TCCAAAAGAA CGTCGTGCGC GTTATGTATC TGACCTTGGT TTATCAGACT ACGATGCTAG 5640

TCAGTTGACT GCTAATAAAG TCACTTCTGA CTTCTTTGAA AAAGCTGTTG CCCTAGGTGG 5700

TGATGCCAAA CAAGTCTCTA ACTGGCTCCA AGGGGAAGTC GCTCAGTTCT TGAATGCTGA 5760

AGGTAAAACA CTGGAACAAA TCGAATTGAC ACCAGAAAAC TTGGTTGAAA TGATTGCCAT 5820

CATCGAAGAC GGTACTATTT CATCTAAGAT TGCCAAGAAA GTCTTTGTCC ATCTAGCTAA 5880

AAATGGCGGT GGCGCGCGTG AATACGTGGA AAAAGCAGGT ATGGTTCAAA TTTCAGATCC 5940

AGCTATCTTG ATCCCAATCA TCCACCAAGT CTTTGCCGAT AACGAAGCTG CTGTTGCCGA 6000

CTTCAAGTCA GGCAAACGTA ACGCCGACAA GGCtTTACAG GATTCCTTAT GAAGGCAACC 6060

AAAGGCCAAG CCAACCCACA AGTTGCCCTT AAACTACTTG CACAGGAATT GGCGAAGTTG 6120

AAAGAAAACT AGACAGAACA AAACCAGCCC TAAGGTTGGT TTTTTCTTCT CTACCAACTC 6180

CCAATAACTA TTTTGGCTTT ATTTCCAGAG TATTTTATGG TAAAATGAAG AGTAATAATA 6240

TTTATTAAAG AGGTAAAAAC ATGATTGAAG CAAGTACCTT AAAAGCTGGT ATGACCTTTG 6300

AAACAGCTGA CGGCAAATTG ATTCGCGTTT TGGAAGCTAG TCACCACAAA CCAGGTAAAG 6360

GAAACACGAT CATGCGTATG AAATTGCGTG ATGTCCGTAC TGGTTCTACA TTTGACACAA 6420

GCTACCGTCC AGAGGAAAAA TTTGAACAAG CTATTATCGA GACTGTCCCA GCTCAATACT 6480

TGTACAAAAT GGATGACACA GCATACTTCA TGAATACAGA AACTTATGAC CAATACGAAA 6540

TCCCTGTAGT CAATGTTGAA AACGAATTGC TTTACATCCT TGAAAACTCT GATGTGAAAA 6600

TCCAATTCTA CGGAACTGAA GTGATCGGTG TCACCGTTCC TACTACTGTT GAGTTGACAG 6660

TTGCTGAAAC TCAACCATCT ATCAAAGGTG CTACTGTTAC AGGTTCTGGT AAACCAGCAA 6720

CGATGGAAAC TGGACTTGTC GTAAACGTTC CAGACTTCAT CGAAGCAGGA CAAAAACTCG 6780

TTATCAACAC TGCAGAAGGA ACTTACGTTT CTCGTGCCTA ATCTCTAGAA AGAGGTCATT 6840

CTATGGGAAT TGAAGAACAA CTTGGCGAAA TCGTTATCGC CCCACGTGTA CTTGAAAAAA 6900

TCATTGCTAT CGCTACTGCA AAGGTAGAGG GTGTTCACTC TTTTTCAAAC AGATCAGTGT 6960

CTGATACCCT TTCAAAACTT TCACTCGGCC GTGGCATTTA TCTTAAAAAC GTGGACGAAG 7020

AACTCACAGC AGATATCTAT CTCTACCTTG AGTACGGAGT AAAAGTTCCT AAGGTAGCGG 7080

TTGCTATCCA GAAAGCTGTC AAAGATGCCG TCCGTAATAT GGCTGATGTA GAACTCGCTG 7140

CTATCAATAT TCACGTTGCA GGTATCGTCC CAGATAAAAC ACCAAAACCA GAATTGAAAG 7200

ATCTATTTGA CGAGGACTTC CTCAATGACT AGTCCACTAT TAGAATCTAG ACGCCAACTC 7260

CGTAAATGCG CTTTTCAAGC TCTCATGAGC CTTGAGTTCG GTACGGATGT CGAAACTGCT 7320

TGTCGTTTCG CCTATACTCA TGATCGTGAA GATACGGATG TACAACTTCC AGCCTTTTTG 7380

ATAGACCTCG TTTCTGGTGT TCAAGCTAAA AAGGAAGAAC TAGATAAGCA AATCACTCAG 7440

CATTTAAAAG CAGGTTGGAC CATTGAACGC TTAACGCTCG TGGAGAGAAA CCTCCTTCGC 7500

TTGGGAGTCT TTGAAATCAC TTCATTTGAC ACTCCTCAGC TGGTTGCTGT TAATGAAGCT 7560

ATCGAGCTTG CAAAGGACTT CTCCGATCAA AAATCTGCCC GTTTTATCAA TGGACTGCTC 7620

AGCCAGTTTG TAACAGAAGA ACAATAAGGC TCTTTGTCAA CTGTAGTGGG TTGAAAAAAA 7680

GCTAAGCTCG AGAAAGGACA AATTTCGTCC TTTCTTTTTT GATGTTCAAA GCGATAAAAA 7740

TCCGTTTTTT GAAGTTTTCA AAGTTTCGAA AACCAAAGGC ATTGCGCTTG ATAAGTTTGA 7800

TGAGATTATT GGTCGCTTCC AGTTTGGCAT TAGAATAGTG TAGTTGAAGG GCGTTGACAA 7860

TCTTTTCTTT ATCTTTGAGG AAGGTTTTAA AGACAGTCTG AAAAATAGGA TGAGCCTGCT 7920

TAAGATTGTC CTCAATAAGT CCGAAAAATT TCTCTGGTTC CTTATTCTGG AAGTGAAACA 7980

GCAAGAGCTG ATAGAGCTGA TAGTGGTGTT TCAAGTCTTG TGAATGGCTC AAAAGCTTGT 8040

CTAAAATCTC TTTATTGGTT AAGTGCATAC GAAAAGTAGG ACGATAAAAT CGCTTATCAC 8100

TCAGTCTACG GCTATCCTGT TGAATGAGTT TCCAGTAGCG CTTGATATCC TTGTATTCAT 8160

GGGATTTTCG ATGAAACTGA TTCATGATTT GGACACGCAC ACGACTCATG GCACGGCTAA 8220

GATGTTGTAC AATGTGAAAG CGATCAAGAA CGATTTTAGC ATTCGGGAGT GAAACAGTCT 8280

GGGAGACTGT TTCAGCCTGA GCCTAGGAAT TTGAAAGCGA AGCTGTTTAG CCAAGTCATA 8340

GTAAGGGCTA AACATATCCA TAGTAATAAT TTTGACGCGA CATCGGACAA CTCTATCGTA 8400

GCGAAGAAAG TGATTTCGAA TGATAGCTTG TGTTCTACCC TCAAGAACAG TGATGATATT 8460

GAGATTGTTA AAATCTTGCG CAATGAAGCT CATCTTTCCC TTTGTAAAAG CATACTCATC 8520

CCAAGACATA ATCTCAGGAA GACAAGAAAA ATCATGTTTA AAGTGAAAAT CATTGAGCTT 8580

ACGAATAACA GTTGAAGTTG AGATGGAAAG CTGATGGGCA ATATCAGTCA TAGAAATCTT 8640

TTCAATCAAC TTTTGAGCAA TCTTTTGGTT GATGATACGA GGGATTTGGT GATTTTTCTT 8700

GACGATAGAA GTTTCAGCGA CCATCATTTT TGAACAGTGA TAGCACTTGA ATCGACGCTT 8760

TCTAAGGAGA ATTCTAGTAG GCATACCAGT CGTTTCAAGA TAAGGAATTT TAGAAGGTTT 8820

TTGAAAGTCA TATTTCTTCA ATTGGTTTCC GCACTCAGGG CAAGATGGGG CGTCGTAGTC 8880

CAGTTTGGCG ATGATTTCCT TGTGTGTATC CTTATTGATG ATGTCTAAAA TCTGGATATT 8940

AGGGTCTTTA ATGTCTAGTA ATTTTGTGAT AAAATGTAAT TGTTCCATAT GAATCTTTCT 9000

AATGAGTTGT TTTGTCGCTT TTCATTATAG GTCATATGGG ACTTTTTTTC TACAATAAAA 9060

TAGGCTCCAT AATATCTATA GGGGATTTAC CCACTACAAA TATTATAGAG CCAACAATAA 9120

AAAGAAAAAG TGTTTGATAG ATATCAAACA CTTTTTTCTT TGCCTCCCAC TATCTAAAAA 9180

AATGATAATA GATATAATTG TAAACAAAAA TCCAGATAGG TTTTGCATGA TTGAGAAAGT 9240

TAAAAAAACT ATGGCAGAGA ATCGTTAATC TCAGATTGTC GGTAGAACGA TAAACAAGGG 9300

CAAAAAAGAA ACCAATCAGA CTATAATATA ATAAACTAAT TGGATCTCTG TGAGATAGTA 9360

200

TCAAATGGCT AATCCCAAAG ATGATAGCAG ATAGGATAAC ATCCAAATAG TACTTGGACT 9420

AGGGAAAGAA GGTATTCATA AAATACCCTC TATCAAGAGT CTCCTCAAAA ACAGGACCGA 9480

TGATTACAGG CAGGACAAAA GATAAGATAG TCGATAAAAA GGTTGGTTGT CCATTTGAAA 9540

AAAGCACGGT AAAATACTCA TCATGAATAT TCCTATGATT AATCAAATGA GCATAGCGTG 9600

CCCAAAAATT ACCGAGAATC TGATAAACCA CATAAGTTGC AAATAAGTAG AAGACAAATG 9660

ACCAGTTCCA GCTCTTTTTC TCAAAGATAA AGAGCATCTT TTTCTTTTTT AACCTCCAAA 9720

TTAATAGAAG GAAACTTCCC ACTAATCCCA TTGTTAAAAT AAGAGAATAG ACATCAGCTC 9780

CTAACCCTAA AATGATCGTC ACATACAATC CAATTGTTTG TGGTAAATAG GTAGATAGTA 9840

AAATAATAAG CAAAAATATT CCAAATTGTC TTAGTTTTTT TGTGTTTCTC ATCGTACTTT 9900

TTTGAAAGAT TACCCTGCTC GGAAGCCGTA CTTCCAAGCA TCTATATAAG AATTAAGTGC 9960

CCCTTGCCTC ATATAGGGAG CAAATTCTCT ATAATATAAC CATCTACTAT ATCCATCTTC 10020

CCAAACAGCA AGACCACCTG AAGTTTGCTC CAAGTCCTCA GTTGAAAGAA CTGTAAATGT 10080

ATTTGTACCT GTCATTGCAA GTACCTTCTT AAAATAGATT GTTGTAGGCT CACATTTATA 10140

GTATATTTCT TTTTTTGTCT ATTTTATAGC CCATCTCCTC AACTGGCAAT TTTTCGACCT 10200

GAATTACATT TTTCCATAAA AAATGAGACC TTTCTAGTCT CATTTAGTCA TTCTTAGTAT 10260

TTTCTAAATC GTTGATAGCG TTCTTCCAGC AACTCTTCTA GCGGTTTTTG TGAAAGTCTA 10320

GCCAGCTCCG TTTGGAGTTC TTTTTTGACA CTCTTAATCA GTTCTTTACT AGAAAGTCCT 10380

ATTTCAGAAA TCACCTTATC CACCACGTCC ATTTCTAACA GTTCATGCGA AGTGATTTTC 10440

ATCAGTTCTG CTGCTTCCAT AGCGCGAGTA CCGTCCTTCC ATAAAATGGA AGCAAAGCCT 10500

TCTGGACTGA GAATGGCATA GATAGAATTT TCCAGCATCC AGACACGGTC CGCGACAGCT 10560

AGAGCCAGAG CCCCGCCTGA ACCACCTTCA CCGATAATAA TGGCGATAAT AGGAACTTTC 10620

AGGTCACTCA TTTCCATGAG ATTGCGAGCG ATAGCTTCCC CTTGACCACG TTCTTCCGCT 10680

CCGACACCAG GATAAGCACC TGCTGTATTG ATAAAGGTCA CAACTGGACG GCCAAATTTC 10740

TCAGCCTGTT TCATCAACCG CAGTGCCTTT CGGTAGCCTT CTGGATGTGG TTGGCCAAAA 10800

TTCCGTTTGA GGTTGTCTTG CAAACTCTTG CCTTTTTGGA TACCAACCAC TGTTACAGCT 10860

TGGTCTCCAA GCCAACCAAT ACCACCAACA ACTGCACCAT CATCACGAAA AGAACGGTCA 10920

CCATGTAATT GGATAAATTC ATCAAAAATG CCTGTCGCAA AGTCCAAGGT TGTCAAGCGA 10980

CTCTGCTCAC GCGCTTCTCT GACTATTTTT GCAATATTCA TCTAGGACTC CCTCCATGCA 11040

ATCTGACTAG GCTAGCAATC GTATCTGGTA AGTCTCTTCT TTTGACAATA GCATCCACAA 11100

AGCCATGTTC TAATAGGAAT TCTGCCTTTT GGAAATCCTC AGGCAAGCTT TCACGAACCG 11160

TATTTTCAAT CACACGACGC CCAGCAAAAC CAACCAAGCT CTGTGGTTCA GCCAGAATGA 11220

TATCGCCTTC CATAGCGAAA GAAGCTGTCA CACCACCAGT CGTTGGATCT GTCAAAATGG 11280

TCAGGTAAAA GAGACCAGCA TTTGAATGGC GTTTAACCGC CGCAGAGATC TTAGCCATCT 11340

GCATGAGACT CATGATTCCT TCCTGCATAC GGGCTCCACC AGAGGCTGTG AATAGGACAA 11400

CTGGCAATTT TTCGACAGTC GCATACTCAA ACAAACGAGT GATTTTTTCA CCTACAACCG 11460

TACCCATAGA AGCCATGATA AAGTTAGAAT CCATAATCCC AAGAGCCACA GTCTGACCTT 11520

TAATAAGAGC AGTTCCTGTC ACAACGGCTT CATGCAGACC TGTTTTTTCA CGCATAGATG 11580

CCAGTTTCTT TTGGTAACCA GGGAAATGCA AGGGATCCTT GCTTTCAATC CCTGTAAACA 11640

ATTCTTTGAA GGTTCCCATA TCAATCGTCA AAGCCAAGCG TTCTTGGGCA GAAATACGAA 11700

AGGTATAGCT ACAGTGCGGA CAGATACGTT CACTTCCCAG ATCCTTCTGA TAGATGGTAT 11760

GCTTACAGCC TGGACACTGG GAAAATAATT CATCTGGAAC CTCTGGCTTA GCTTGAGGTT 11820

TTTCCCTAAC CGAACGATTG GGATTGATTC GAATATACTT ATCTTTTTTA CTAAATAGAG 11880

CCATTGATTC CCCTTTTCGG TTTAAACTCT TAAAGTCATT TTATTCTTTT TCTTGATATT 11940

TAGGTAAGAA GGTTTCCATC AAGAAGGAAG TATCATAATC CCCAGCAATG ACATTGCGAT 12000

CTGAAATGAG GTCAAGCTGG AAATCTGCAT TGGTCTGCAC TCCTTCAATT TCTAATTCAT 12060

AGAGGGCACG TTGCATTTTC ATCAAGGCGT CAAAACGATT TTCGCCGTGT ACTATGATTT 12120

TGGCAATCAT ACTATCATAA TAAGGCGGAA TGGTATAACC TGGATAAACT GCTGAATCCA 12180

CGCGCAAGCC AACTCCACCA CTTGGCAGAT AGAGATTAGT AATCTTACCT GGACTTGGAG 12240

CAAAGTTAAA GGCTGGGTTT TCTGCATTGA TACGACACTC GATGGCATGA CCGCGTAGGA 12300

CAATATCTTC TTGCTTAACA GACAAAGGCT GACCTGCCGC AATGCAAATC TGTTCCTTAA 12360

CGATATCAAC ACCTGAAACA AACTCTGTTA CTGGATGTTC TACCTGAACA CGAGTATTCA 12420

TCTCCATGAA ATAGAAATTG CTACTTGCTT CATCAAGAAG AAATTCAATG GTTCCTGCAT 12480

TCTCATAGCC AACAAACTCT GCCGCTCGAA CAGCAGCAGC ACCTATTTCA TGACGCAGCG 12540

TTTTTCCGAT TGCAATCGAG GGACTTTCTT CCAAAACCTT TTGGTTATTC CTTTGAAGAG 12600

AACAATCCCG TTCACCCAAG TGAATCACAT GTCCATGCTC ATCACCTAGG ATTTGAACCT 12660

CAATGTGCCG AGCTGGATAG ATAACCCGTT CTATGTACAT GGCACCATTG CCATAATTGG 12720

CCTTGGCCTC ACTAGAGGCA GTTTCAAAGG CAGAAACGAG GTCATCTGGT TTTTCAACCT 12780

TACGAATCCC TTTACCACCT CCACCTGCTG AAGCCTTGAG CATAACAGGA TAGCCAATTT 12840

TTTCAGCAAC AATCAAAGCT TCTTCAGAGT TATGCACTTC TCCATCTGAA CCTGGTATAA 12900

CAGGCACACC TGCTTTAATC ATCTGAGCAC GCGCATTGAT CTTATCCCCC ATCATATCCA 12960

TAACATGACC AGATGGACCG ATAAACTTGA TACCTACTTC TTCACACATG GTCGCAAATT 13020

TGGAATTTTC ACTGAGAAAT CCAAAACCAG GGTGAATAGC TTCTGCCTCA GTCAAGACTG 13080

CAGCTGATAG AACTGCATTA ATATTGAGAT AAGACTCTGT TGCCTTGCCA GGACCAATAC 13140

AAACTGCTTC ATCTGCCAAA AGCGTATGAA GAGCTTCCTT ATCAGCAGTT GAATAAACCG 13200

CTACCGTCGC AATCCCCAAT TCACGTGCCG CACGGATAAT ACGAACCGCA ATTTCACCAC 13260

GATTGGCAAT TAAAATTTTT CGAAACATGG AGAACCTCCT TAGTTCCCAA TTGCAAAAGT 13320

AAGGGTACCA CTGGCTGCAA GCTTGCCATC CACTTCAGCC TTTGCTTCAA CCACAGCTAT 13380

GGTGCCACGA CGTTTTACAA AAGTCGCTGT CATAACCAAT TGGTCGCCTG GTACAACTTG 13440

CTTCTTGAAC TTAACCTTGT CCATACCAGC GTAAAAGACC AGTTTTCCTT TATTTTCAGG 13500

TTTTGATAAC TCCAACACAC CGGCAGTTTG CGCCAAGGCT TCCATAATCA CAACACCTGG 13560

CATAACTGGG TATTGAGGAA AGTGGCCGTT AAAGAAAGGC TCGTTGATGG TCACATTTTT 13620

GATAGCAACA ATGGTATCCT CGCTCACTTC CAAGACACGG TCCACTAGAA GCATAGGATA 13680

ACGGTGGGGA AGAGCTTCTT TGATTCCTTG AATATCGATC ATTTGATACG TACCAATCCT 13740

TTACCAAACT CAACCATTTC TTCGTTAGAG ACGAGAATTT CCGTTACCAC ACCATCCTTA 13800

GGAGCTGGGA TTTCATTCAT GACTTTCATG GCTTCGATAA TTACCAATGT TTGACCTTTT 13860

TTGACACTAT CACCAACTGT AACGAAGGCA GGTTTATCTG GTCCAGCAGC CAAGTAAACC 13920

ACTCCAACAA GTGGACTCTC TACAAGATTT CCCTCAGTAG CCACACTTGC TTCAGCTGGA 13980

GCTGGAACTT CTTCTGCTAC AGTCTCTGCT GGAGCAGATG TAGGAGCTAC TGGACTCGGT 14040

GTTGCTAGAA CGGGTGCTGG AGCGACTTGA GTTGCAACTT CAGGCACAGG TCTTGCTTCA 14100

TTCTTGCTAA ACTGCAACTC ATCCGTCCCA TTTTTATAAG AAAATTCTCT CAAACTTGAC 14160

TGGTCAAATT GAGTCATCAA GTCTTTAATA TCGTTTAAAT TCATACTTAT CTATTCTCCC 14220

AACGTTTGAA AGCAAGAACT GCATTGTGGC CTCCAAAACC AAAAGTATTT GAAATAGCGT 14280

ATGGAATTTC TTTCTCCAAG CCTTGTCCAT AAACGACATT AGCTTCGATA TAATCTGATA 14340

CTTCACTTGT CCCAGCTGTC ATTGGTACAA AGTTATGACG CATAGCTTCG ATGGTGACGA 14400

TAGCTTCTAC TGCACCCGCA GCCCCCAGCA AATGTCCTGT AAAAGACTTG GTTGATGATA 14460

CAGGTACTTC CTTACCAAGA ACAGCTACGA TAGCACCACT TTCTCCTTTT TCATTGGCAG 14520

GAGTTGACGT TCCGTGAGCA TTGACATAGG CTACTTGCTC TGGAGAAATC TCAGCTTCTT 14580

CCAAGGCTAG TTTGATGGCC TTGATAGCTC CCTGACCTTC TGGATGTGGA GAAGTCATGT 14640

GGTAGGCATC ACAAGTATTT CCGTAACCAA CCACTTCAGC CAGGATAGTA GCTCCACGTT 14700

TTTCAGCGTG TTCAAGACTT TCTAGAACCA ACATCCCTGA ACCTTCACCC ATAACAAACC 14760

CATTGCGATC CTTATCAAAT GGGATCGAAG CACGAGTTGG ATCCTCTGTA GTAGAGAGAG 14820

CTGTTAAGGC TTGGAAACCA GCGATGGCAA AAGGTGTGAT AGAAGCTTCT GTTCCTCCCA 14880

CCAACATCAC ATCTTGGAAA CCAAACTTAA TGGAGCGGAA GGCATCCCCA ATCGCATCAT 14940

TTGATGAAGA GCAGGCAGTA TTGATAGATT TACAAACACC GTTTGCACCA AAACGCATGG 15000

CTACATTCCC AGAAGCCATA TTTGGTAAAG CTTTTGGAAG AGTCATTGGT TTGACACGTT 15060

TGGGTCCTTT TTCATGAAGG CGAAGTACCT GATCTTCAAT TTCCTTGATT CCACCAATAC 15120

CAGATGCAAC GATAACACCA AAACGATCCC TATTAAGAGC CTCTACATCA AGATTGGCAT 15180

GATTTACAGC CTCTTGGGCT GCATACAAGG CATATAAAGA ATAGTTATCA AAACGGTTGG 15240

TATCTTTTTT TACAAAGTAT TTATCGAACG GAAAATCTTG GATTTCTGCC GCATTATGCA 15300

CATCAAAGTC ACTATGATCA AATTTTGTAA TGCCACCAAT GCCGATTTTC CCAGTTGCTA 15360

AACTATTCCA AAATTCTTCT GGTGTATTTC CGATTGGAGA TGTTACTCCA TAACCTGTTA 15420

CCACTACTCG ATTTAGTTTC ATTCTTTTCA CCTCTAGCTT TCGCTACATA CTTAAGCCAC 15480

CATCAATGGC AACCACTTGT CCAGTTAGAT AATCTTGGCC TGCTAAAAAT ACTGTCAAAT 15540

CTGCAACCTG CTCTGCCTGC CCAAATTCTT TCATCGGAAT CTGAGCTAGT GTAGCTTCCT 15600

TAATCTTATC TGACAGGATA GCGGTCATAT CAGACTCAAT CATTCCTGGA GCAATCACAT 15660

TGACTCGTAT ATTCCGACTA GCGACCTCGC GTGCCACAGA CTTGGTAAAG CCAATCAAGC 15720

CAGCCTTAGA AGCAGCATAA TTAGCTTGAC CAATATTCCC CATCAAACCA ACAACACTAG 15780

ACATATTAAT GATAGCACCT TCTCTGGCTT TCATCATCGG TTTCAAGACT GATTGTGTCA 15840

TATTAAAGGC ACCAGTCAGA TTGACCTTGA GCACTTTTTC AAAATCTGCT TCTGTCATCT 15900

TGAGCATAAG AGTATCTTGG GTAATCCCTG CATTGTTGAC CAAAACATCT ACTGAACCCA 15960

GTTCTGCAAT AGCTTGATCA ATCATACGCT TAGCGTCTGC AAAATCTGAT ACATCTCCTG 16020

AAATGGGAAC CACCTTGATA CCATAGTTTG AAAACTCAGC GAGCAATTCT TCTGAGATTG 16080

CCCCACGACT GTTTAAGACA ATGTTGGCTC CTGCTTGAGC AAACTTGTGG GCGATGGCAA 16140

GACCAATTCC ACGACTCGAA CCTGTAATAA AGATATTTTT ATGTTCTAGT TTCATTTTTT 16200

TCCTTTCAAA ACTTCTACTT ATTTTAGTCT ATTTTTCTAA AAGTGCTACT AAACTCGCTT 16260

GATCTTCCAC ATGAGCTAAG TGAGCAGTTT GATCAATTTT TTTAACAAAA CCTGACAAGA 16320

CTTTCCCCGG TCCAATCTCG ATAAAGTTGC TTATGCCTGC TTCTTGCATG ACCCCAATAC 16380

TTTCATAGAA ACGAACGGGT TCCTTGACCT GACGCGTCAA GAGCTGAGCA ATGTCCTCTT 16440

TTTGCATCAC AGCAGCTTCT GTATTGCCGA CTAGGGGACA AGTAAAATCT GAAAAACTTA 16500

CCTGAGCTAG AGTTTCAGCT AGTTTCTGGC TAGCAGGTTC AAGGAGAGCG GTGTGAAAGG 16560

GACCTGACAC CTTAAGAGGA ATCAAGCGTT TGGCACCTGC TTCTTGCAAA AGTTCAACCG 16620

CTCGATCAAC TGCAACCACT TCTCCAGCAA TGACGATTTG TGCAGGTGTG TTATAGTTGG 16680

CTGGAGTAAC CACTCCAAGT TCAGAAGCTT TTTGACAGGC TTCTTCAATG ACCTCTACTG 16740

GCGTATTGAG AACTGCTACC ATCTTGCCAG AGTCAGCAGG AGCCGCTTCT TCCATATAGG 16800

CTCCACGCTT AGCTACCAAG GCAACCGCAT CTTCAAAATC CAAGGCGCCA CTTGCCACCA 16860

AGGCAGAGTA TTCTCCAAGA GACAAACCAG CAACCATATC AGGCTGATAG CCCTTTTCTT 16920

GCAATAAACG GTAGATAGCA ACCGAAGTCG CTAGAATGGC TGGTTGCGTA TAGCGGGTCT 16980

GATTGAGTTT GTCTTCTTCC GTATCGATGA GATAACGCAA ATCATAACCG AGCACCTGGC 17040

TCGCTCGATC AATCGTTTCT TTAACAATCG GATACTGATC ATAGAAATCC CGTCCCATCC 17100

CTAGATACTG GGCACCTTGA CCAGCAAATA AAAAGGCTGT TTTAGTCATT TCTTACAACT 17160

CCTGTCCAGC GAGAGGCTTC TTCTTGAATT TTCTTAGCGG CTCCGTAATA CAAATCTTTT 17220

AGGATTTCTT CAGCTGTTTC TTCTTTAGAA ACAAGCCCTG CGATTTGACC TGCCATAACA 17280

GAGCCACCAT CCACATCACC GTGAACAACT GCTTTGGCTA GAGCACCTGC TCCCATTTGT 17340

TCAAAGATTT CTAAATCAGG ATCTTCTTGC TTAAAGGCAT CTTTTTCAGC CAGTTCAAAA 17400

TCTCTAGTCA ACTGATTTTT AATAGCACGA ACAGCATGAC CAAAGTGCTG AGCTGAAATC 17460

GTAGTATCAA TATCCCTTGC TTTTAAAATT TTCTCCTTGT AGTTTGGATG GGCATTCGAC 17520

TCTTTTGCAA CTACAAACCG TGTCCCCACC TGTACAGCCT CTGCACCTAG CATAAAGCCA 17580

GCCGCAGCAC CTTCACCATC CGCAATTCCT CCTGCAGCAA TAACAGGAAT AGATATAGCT 17640

GTGGCTACCT GTCGCACCAA GGTCATGGTT GTTAATTTAC CGATATGCCC CCCAGCTTCC 17700

ATTCCTTCTG CAATAACAGC GTCTGCACCG ATTTTTTCCA TGCGTTTAGC TAAAGCGACA 17760

CTAGGAACAA CAGGAATAAC GATTATCCCA GCTTCATGGA AACGTTCCAT ATACTTGCTT 17820

GGATTTCCTG CTCCTGTTGT GACAACTTTA ACACCTTCTT CAATAACGAG ATCCACGATG 17880

TCTTCCACAA AGGGAGATAA GAGCATGATG TTGACCCCAA AGGGTTTATC AGTCAATGAT 17940

TTGATTTTAT CAATATTGGC CTTGACAACT TCTTTCGGGG CATTTCCCCC ACCGATAATT 18000

CCTAATCCTC CAGCCTTGGA AACAGCCCCT GCCAAATCAC CATCAGCAAC CCAGGCCATC 18060

CCTCCTTGGA AAATAGGATA ATCAATCTTC AATAATTCTG TAATACGCGT TTTCATAGTG 18120

CCTCCAACCT TCCTTGCTTA CGTAATAGTT CGATTTCACC ATAATTTGAC AGTCAAACTA 18180

TTACCTAAAC AAGAGGGAGT GGGTTTCTCC CTACTCCTTC TACTAATATT CTGCTTATTT 18240

TGCTTGCTCT TCAACGTAAG CAACCAAGTC ACCAACTGTT TTCAAGTCAT TTTCTGCTTC 18300

GATTTGGATA TCAAAAGCAT CTTCGATTTC TGAGATTACT TGGAACAAGT CCAATGAATC 18360

TGCGTCCAAA TCATCAAAAG TTGATTCAAG TGTTACTTCT GATGCGTCTT TTCCAAGTTC 18420

TTCAACGATA ATTTCTTGTA CTTTTTCAAA TACTGCCATG ATAGGACTCC TTTAAAATAA 18480

ATAGTTTTTT TATAACAATG TGTTCACCAC ATGATTACCT AAATTGTAAG AATGAGCGTG 18540

CCCCAGGTCA AGCCTCCACC GAAGCCTGAT AGAAGAACAG TCTGGCTACC ATCTAAAGGG 18600

ATGAGACCTT GTTCTACACA CTCTGAAAGT AAAATCGGGA TACTGGCTGC ACTGGTATTG 18660

CCATATTCCA TCATATTGGC TGGAAGTTTG GCTCGGTCAA CACCAATTTT TCTAGCCATC 18720

TTATCCAAAA TACGGTCATT GGCTTGATGA AGTAGCAGAT AATCCAAGTC TGTCACCTCT 18780

ATAGGAGATT CATCAATAGT CTGCTTGATA GACTTGGCTA CATCTCGAAT GGCAAAATCA 18840

AAGACTGTGC GTCCATCCAT CTTCAAAAAC GAATCTGCAC TTTCTTGATC TGAAAATGGA 18900

GAATGTAAAC CTGAATGCCC ATAAGTTAAA CACTCGCTGC GACTTCCATC GCTATTGAGA 18960

CTCTCAGCTA AGAAATGCTC TTGCTCGCTA GCTTCTAACA AGACACCACC AGCACCATCT 19020

CCAAACAACA CAGCTGTTGA TCGATCCGAC CAATCGACTG CCTTAGAGAG GGTTTCACTA 19080

CCAATCACCA AGCCTTTTTG AAAGCGACCA GAAGCGATAA ACTTTTCAGC AGTTGAAAGA 19140

GCAAATACAA ATCCACTGCA AGCCGCGGTT AAGTCAAAAG CAAAGGCTTT ATTAGCACCA 19200

ATATTAGCTT GAACACGAGC AGCTGTAGAG GGCATCATCG AATCTGGAGT AATGGTAGCT 19260

AGGATGATAA AATCCAGTTC TTCTCCTGTT ATTCCAGCTT TTGCCATCAG TTTCTTAGCA 19320

ACCTCTGTAG CCAAATCACT GGTAGATTCT GTTCTTGAAA TATGCCTTTG TCGTATTCCC 19380

GTTCGACTTG AAATCCACTC ATCATTGGTA TCCATAATCT GAGCCAAGTC GTGATTTGTA 19440

ACCACTTGCT CTGGCACATA ATGAGCAACC TGACTTATTT TTGCAAAAGC CATTATTTCA 19500

AATCCTCCAA AAATTGGTAA AGATTAGTCA AACCTTTACC CATGACAGCA ATTTCTTCCT 19560

CGCTCATGCC ATCAATAATT TTTTCTACCA TGGCCTTGTG GAAGCGTTTA TGCAGTCTAT 19620

GAATCAAGCG ACCCTTCTTT GTCAAATGCA GATGCACCAC ACGACGATCC TGTTCTGACC 19680

GAACTCGCTC AATGTAGCCC GG 19702

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5211 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GAAAATTTCC TCTCTTCTCT TGAAAAATTT TGAAAAAATG GTATGATAGT AACAAGTTAT 60

TTTTAAGAGG AAAGAAAGGG GAATAATGGA GAAAATCAGT TTAGAATCTC CTAAGACGGG 120

GTCGGACCTA GTTTTGGAAA CACTTCGTGA TTTAGGAGTT GATACCATCT TTGGTTATCC 180

TGGTGGTGCG GTTTTGCCTT TTTATGATGC GATATATAAT TTTAAAGGCA TTCGCCACAT 240

TCTAGGGCGC CATGAGCAAG GTTGTTTGCA TGAAGCTGAA GGTTATGCCA AATCAACTGG 300

AAAGTTGGGT GTTGCCGTCG TCACTAGTGG ACCAGGAGCA ACAAATGCCA TTACAGGGAT 360

TGCGGATGCC ATGAGCGATA GCGTTCCCCT TTTGGTCTTT ACAGGTCAGG TGGCGCGAGC 420

AGGGATTGGG AAGGATGCCT TTCAGGAGGC AGACATCGTG GGAATTACCA TGCCAATCAC 480

TAAGTACAAT TACCAAGTTC GTGAGACAGC TGATATTCCG CGTATCATTA CGGAAGCTGT 540

CCATATCGCA ACTACAGGCC GTCCAGGGCC AGTTGTAATT GACCTACCAA AAGACATATC 600

TGCTTTAGAA ACAGACTTCA TTTATTCACC AGAAGTGAAT TTACCAAGTT ATCAGCCGAC 660

TCTTGAGCCG AATGATATGC AAATCAAGAA AATCTTGAAG CAATTGTCCA AGGCTAAAAA 720

GCCAGTCTTG TTAGCTGGTG GTGGAATTAG TTATGCTGAG GCTGCTACGG AACTAAATGA 780

ATTTGCAGAA CGCTATCAAA TTCCAGTGGT AACCAGTCTT TTGGGACAAG GAACGATTGC 840

AACGAGTCAC CCACTCTTTC TTGGAATGGG AGGCATGCAC GGGTCATTCG CAGCAAATAT 900

TGCTATGACG GAAGCGGACT TTATGATTAG TATTGGTTCT CGTTTCGATG ACCGTTTGAC 960

GGGGAATCCT AAGACTTTCG CTAAGAATGC TAAGGTTGCC CACATTGATA TTGACCCAGC 1020

TGAGATTGGC AAGATTATCA GTGCAGACAT TCCTGTAGTT GGAGATGCTA AGAAGGCCTT 1080

GCAAATGTTG CTAGCAGAAC CAACAGTTCA CAACAACACT GAAAAGTGGA TTGAGAAAGT 1140

CACTAAAGAC AAGAATCGTG TTCGTTCTTA TGATAAGAAA GAGCGTGTGG TTCAACCGCA 1200

AGCAGTTATT GAACGAATTG GTGAATTGAC GAATGGAGAT GCCATTGTGG TAACAGACCT 1260

TGGTCAACAC CAAATGTGGA CAGCTCAGTA TTATCCCTAC CAAAATGAAC GTCAGTTAGT 1320

GACTTCAGGT GGTTTGGGAA CAATGGGCTT TGGAATTCCA GCAGCAATCG GTGCTAAAAT 1380

TGCTAACCCA GATAAGGAAG TAGTCTTGTT TGTTGGGGAT GGTGGTTTCC AAATGACCAA 1440

CCAGGAGTTG GCTATTTTGA ATATTTACAA GGTGCCAATC AAGGTGGTTA TGCTGAACAA 1500

TCATTCACTT GGAATGGTTC GCCAGTGGCA GGAATCCTTC TATGAAGGCA GAACATCAGA 1560

GTCGGTCTTT GATACCCTTC CTGATTTCCA ATTGATGGCG CAGGCTTATG GTATTAAAAA 1620

CTATAAGTTT GACAATCCTG AGACCTTGGC TCAAGACCTT GAAGTCATCA CTGAGGATGT 1680

TCCTATGCTA ATTGAGGTAG ATATTTCTCG TAAGGAACAG GTGTTACCAA TGGTACCGGC 1740

TGGTAAGAGT AATCATGAGA TGTTGGGGGT GCAGTTCCAT GCGTAGAATG TTAACAGCAA 1800

AACTACAAAA TCGTTCAGGA GTCCTCAATC GCTTTACAGG TGTCCTATCT CGTCGTCAGG 1860

TTAATATTGA AAGCATCTCT GTTGGAGCAA CAGAAGATCC GAATGTATCG CGTATCACTA 1920

TTATTATTGA TGTTGCTTCT CATGATGAAG TGGAGCAAAT CATCAAACAG CTCAATCGTC 1980

AGATTGATGT GATTCGCATT CGAGATATTA CAGACAAGCC TCATTTGGAG CGCGAGGTGA 2040

TTTTGGTTAA GATGTCAGCG CCAGCTGAGA AGAGAGCTGA GATTTTAGCG ATTATTCAAC 2100

CTTTCCGTGC AACAGTAGTA GACGTAGCGC CAAGCTCGAT TACCATTCAG ATGACGGGAA 2160

ATGCAGAAAA GAGCGAAGCC CTATTGCGAG TCATTCGCCC ATACGGTATT CGCAATATTG 2220

CTCGAACGGG TGCAACTGGA TTTACCCGCG ATTAAAAATC CAACTTAAAT TTATTAAACC 2280

AGCCTAAAAG GCAATAAATA ATAGAAAAGA GAGAAAAGCT ATGACAGTTC AAATGGAATA 2340

TGAAAAAGAT GTTAAAGTAG CAGCACTTGA CGGTAAAAAA ATCGCCGTTA TCGGTTATGG 2400

TTCACAAGGG CATGCGCATG CTCAAAACTT GCGTGATTCA GGTCGTGACG TTATTATCGG 2460

TGTACGTCCA GGTAAATCTT TTGATAAAGC AAAAGAAGAT GGATTTGATA CTTACACAGT 2520

AGCAGAAGCT ACTAAGTTGG CTGATGTTAT CATGATCTTG GCGCCAGACG AAATTCAACA 2580

AGAATTGTAC GAAGCAGAAA TCGCTCCAAA CTTGGAAGCT GGAAACGCAG TTGGATTTGC 2640

CCATGGTTTC AACATCCACT TTGAATTTAT CAAAGTTCCT GCGGATGTAG ATGTCTTCAT 2700

GTGTGCTCCT AAAGGACCAG GACACTTGGT ACGTCGTACT TACGAAGAAG GATTTGGTGT 2760

TCCAGCTCTT TATGCAGTAT ACCAAGATGC AACAGGAAAT GCTAAAAACA TTGCTATGGA 2820

CTGGTGTAAA GGTGTTGGAG CGGCTCGTGT AGGTCTTCTT GAAACAACTT ACAAAGAAGA 2880

AACTGAAGAA GATTTGTTTG GTGAACAAGC TGTACTTTGT GGTGGTTTGA CTGCCCTTAT 2940

CGAAGCAGGT TTCGAAGTCT TGACAGAAGC AGGTTACGCT CCAGAATTGG CTTACTTTGA 3000

AGTTCTTCAC GAAATGAAAT TGATCGTTGA CTTGATCTAC GAAGGTGGAT TCAAGAAAAT 3060

GCGTCAATCT ATTTCAAACA CTGCTGAATA CGGTGACTAT GTATCAGGTC CACGTGTAAT 3120

CACTGAACAA GTTAAAGAAA ATATGAAGGC TGTCTTGGCA GACATCCAAA ATGGTAAATT 3180

TGCAAATGAC TTTGTAAATG ACTATAAAGC TGGACGTCCA AAATTGACTG CTTACCGTGA 3240

ACAAGCAGCT AACCTTGAAA TTGAAAAAGT TGGTGCAGAA TTGCGTAAAG CAATGCCATT 3300

CGTTGGTAAA AACGACGATG ATGCATTCAA AATCTATAAC TAATTAGAAA TATATAGCGC 3360

TGGAGATGAT TTTATGAAAA AGATTATGAG AAAAATTGCA TCGTTATTAT TGGTTCTAGT 3420

TGTATAATGT AATTACACCG TCGGTAATAG TGCTAGCAGA CCAAAATAAA GCAGATTGGT 3480

CGTATGATGA AAATGCTGTA ATTAACATTT ATGATGATGC TAATTTTGAA GATGGTAGGT 3540

TGCATATGAA CTTTGAACAA TTCTTCAAAT TGGCACAAAT AGCTAGAGAA GAAGGTCTTG 3600

AAATTCATTC TCCGTTTGAG AGAGCTGGTG CGACTAAATC TGCTCGTTAT ATAGCGAAAT 3660

GGATTTTGAG AAATAAAAAA CATTAACAAA TATAGTTGGT AAATCATTAG GACCTAAATC 3720

AGCTGTTAGA TTCGGAGAAG CTTTATCCTA TATTGAAGGT CCTCTTCGCA GAATAAATGA 3780

GACGATAGAT GGCGGTTTAT ATCAAATAGA GCAAATTATT GCATCTGGAT TGAAAGAATC 3840

GGGTTTAAAT GACTGGACTG CGAAAACTTT AGCTTCAGCT ATTCGTGGGA TATTAGATGT 3900

ACTTATTTAG GGGTTGAAAT CATATGAATA TTACCAATTT GTTTTCTATC AAGACAGGAT 3960

GTGATGAAAC TGATAGGCAA CTGCAAAAAC TATTTTTTCA GTTGGATTTA CAATTGGGAG 4020

AATTGACAGA TCAACTAAGA AAATTAGATT CTAATTTTGT TCCTCGTAGT CAATTTGTAG 4080

ACACGTTGGA TTTGAATGAT GTAGAATATA AAGAAATTTT AAACTATTTT ATCTTCCATC 4140

GTAATGATAG TGAAGAAAGT TTGGTAGAAT GGTTATATGA TTGGATTTCC ACAAATCGTT 4200

ATGAACTTCC TAAAGAGTTT TCGATTCGTA TGGCTCATAA ATACCATGAA AGTGTTACTG 4260

AAGTTTTCGG AGATGAATAA CTAAAAAACA GTCATTAGTG ACTGTTTTTT ATAGAAAAAG 4320

AGGTTTTATA TGTTAAGTTC AAAAGATATA ATCAAGGCTC ACAAGGTCTT GAACGGTGTG 4380

GTTGTGAATA CTCCACTGGA TTACGATCAT TATTTATCGG AGAAGTATGG TGCTAAGATT 4440

TATTTGAAAA AAGAAAATGC CCAGCGTGTT CGCTCCTTTA AAATTCGTGG TGCCTATTAT 4500

GCCATTTCCC AGCTCAGCAA GGAAGAACGT GAACGTGGGG TAGTCTGCGC TTCTGCGGGA 4560

AATCATGCGC AGGGAGTAGC CTATACTTGT AATGAAATGA AAATTCCTGC TACTATCTTT 4620

ATGCCCATTA CTACGCCACA ACAAAAGATT GGTCAGGTTC GCTTTTTTGG TGGGGATTTT 4680

GTAACTATTA AACTAGTTGG AGATACCTTT GATGCCTCAG CCAAAGCAGC TCAAGAATTT 4740

ACAGTCTCTG AAAATCGTAC CTTTATTGAT CCTTTTGATG ATGCTCATGT TCAAGCAGGT 4800

CAAGGAACAG TTGCTTATGA GATTTTAGAA GAAGCTCGAA AAGAATCGAT TGATTTTGAT 4860

GCTGTCTTGG TTCCTGTTGG TGGTGGCGGT CTCATTGCCG GGGTTTCTAC CTATATCAAG 4920

GAAACAAGTC CAGAGATTGA GGTTATCGGA GTAGAGGCGA ATGGAGCGCG TTCCATGAAA 4980

GCTGCCTTTG AGGCTGGAGG TCCAGTAAAA CTCAAGGAAA TTGATAAATT TGCTGATGGG 5040

ATTGCTGTGC AAAAGGTAGG TCAGTTGACC TATGAAGCAA CTCGTCAACA TATTAAAACT 5100

TTGGTAGGTG TCGATGAGGG ATTGATTTCT GAAACCTTGA TTGACCTTTA CTCTAAGCAA 5160

GGGATAGTCG CAGAACCTGC TGGAGCGGCT AGTATCGCCT CTTTAGAGGT TTTAGCTGAA 5220

TATATTAAGG GGAAAACCAT TTGTTGTATC ATTTCTGGAG GAAATAATGA TATCAACCGT 5280

ATGCCAGAAA TGGAAGAGCG TGCCTTGATT TATGATGGTA TCAAACATTA CTTTGTGGTC 5340

AATTTCCCAC AACGTCCAGG AGCTTTGCGT GAGTTTGTAA ATGATATCCT GGGGCCAAAT 5400

GATGATATCA CACGTTTTGA GTATATCAAA CGAGCTAGCA AGGGAACAGG CCCAGTATTA 5460

ATTGGGATCG CTTTAGCAGA TAAGCATGAT TATGCAGGTT TGATTCGTAG AATGGAAGGT 5520

TTTGATCCAG CTTATATTAA CTTAAATGGT AATGAAACGC TTTATAATAT GCTTGTCTGA 5580

GGACTAATAA AAAAATATCA TACCTTCATT TTGATTTCCT ATCTATTGAC AAGCATAGTC 5640

ACACTGTCTT TAATACTCTT CGAAAATCTC TTCAAACCAC GTTAGCTCTA TCTGCAACCT 5700

CAAAACAGTG TTTTGAGCAA CTTGCGGCTA GCTTCCTAGT TTGCTCTTTG ATTTTCATTG 5760

AGTATAAGGT ATGATTTGAT TTCTTTTTGT TGACAAATAT ACTATATTAA AAAGATATAT 5820

AAGTAATTAA CTGAGCTTAT CTGTCTTGTC ATCTCTATTA AGGATGGTTT AGATAATCGG 5880

GTGTCTGCTT CTAGGCTAGC ACCTCAATAT CCAAAGGAGT GATGAATTTG AAGGACATAA 5940

GGAATACCTA TCTCTCAGAT GATTTATTGA GGAAGAAAGA TAGGAGTTTT TGAGCTAGTG 6000

AAGGCTTGGA TTTCTAAAGG TTAGAACTAT CATCTTCAGT TCTTAAATCG AAGAAATAAG 6060

CTATCTTACG GAAATAGAGA AGCATTTTTT AAGAACTTGA ATAATTTCGC ACCTTAAGAG 6120

GGTAATAATA CAGTATTTTT ATTAGCAAAT ATTTATGGTG TAGAGGCTAG CAAAACCTAT 6180

ATATTATCGG ATTTAAAAAG GAAGTAAGAA A 6211

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7939 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CCGGACTCCC CACGATTCTT CAAAATAACT GAGTATATTT CTATCTTGAT TTTCAGATAT 60

AAATTCTTCC TTCTGTGGCC TCTTCTTACG CTTGAGAAGA GCTTCTCCGA CATGGCTTCT 120

TCCTTACTGA GCAAAACCTT GAGCATAGAT AAGTTTGACT GGCAAGCGTG CTCTTGTATA 180

TTTGGCTCCC TTCCCACTAT TGTGGATAGC GAGGCGTCTT CTCATATCAG TCGTATAGCC 240

TATATAGTAG GATCCATCAC GACACTCCAG AACGTACATA TAAGCCTTAT GATCCATAAT 300

AAATCTCTTC GATTTCGGGC GTATAAGAGC CATCATCATT GTGGACAATC AAAGGAGGTA 360

AGACCTTAAA GCCACTTGTT GAGCCATCCT TGATCGCCTC AATCAAAAGC ATATTGGCTT 420

CCTTTTCTCT TTTTGGATAA ACAAACTGCA GGCGCTTAGG GGCTAGATTA TGTCGTTTTA 480

ACGTATCCAA AATATCCAGA AGTCGATCAG GACGATGAAC CATGGCCAAA CGCCCATTAG 540

ACTTGAGAAT ACTCTGGGCA CTACGACAGA TTTCTTCCAA ATTAGTCGTG ATTTCGTGTC 600

GAGCCAAGAG ATAATGTTCA CTCTCGTTCA GATTAGAATA AGGATTCACC TTGAAATAGG 660

GTGGATTACA CAAAATCATA TCCACCTTAC TCCCCTGAAT GTGAGCAGGC ATATTTTTCA 720

AATCATCGCA GATGACCTGC ATTTGCTCCT CTAATCCATT CAAACGGACA GAGCGTTCAG 780

CCATATCCGC CAAACGCTCC TGAATCTCAA CAGACAATAT CTGTGCTTGA GTACGAGTGC 840

TAGCAAAAAG CCCCACTGCT CCATTCCCAG CACAGAAATC CACAATCAAC CCCTTCTTAG 900

GAAAACGTGG AAATCGTGAT AAGAGAACAC TATCCACCGA ATAGCTAAAA ACCTCTCTAT 960

TTTGAATGAT TTTGATATCT GTCGAAAAGA GCTGGTTAAT GCGCTCTCCT GATTTTAATA 1020

ATTGTTCTTC TTCCATGGTC CTATTATAGC AAATTCATAT TAACATTACA AAAAATATAA 1080

AACTCTAAAC TACTTCTTCT TTTTTAAATG GTGCAGGGCT TCTCCAGTCC AGATTGGTAG 1140

CATTCGTCGA AAGGGAGCAA AGCCGTAGTT AAAGCGGTCG CTTGAAAAGC GTCTCCGTCT 1200

AGGAAACTGG TACTTTTCTT CCTCCAAAGT GCGGATAGAA AGACTGGCTT TCCCTGTAAA 1260

TTCATCTAAA TCCACTACCT GAACTTGAAC CTCTTCATCG ACTTTCAAGG TTTCATGAAT 1320

ATTTTCAATA AATCCTGTCC GAATCTCTGA AATGTGAATC AGCCCCGTAT CACCCGTCTC 1380

TAACTCAACA AAGGCACCGT AGGGCTGAAT CCCTGTAATA CGCCCCTTTA GCTTATCACC 1440

GATTTTCATC TTAGTCCTCG ATTTCAATAG TTTCAATTAC AACATCTTCA ACTGGCTTGT 1500

CCATAGCTCC TGTCTCAACA GCAGCAATGG CATCCAAGAC AGCGTAAGAT GCTTCATCAG 1560

CTAACTGACC AAAAACCGTG TGACGGCGGT CTAGGTGAGG TGTCCCACCT TGATTGGCAT 1620

AGATTTCTGC AATCGGTTCT GGCCAACCAC CACGAGTAAT TTCTTTCTTA GAATAAGGTA 1680

GGTGTTGGTT TTGCACGATA AAGAACTGGC TGCCGTTGGT ATTTGGACCA GCATTTGCCA 1740

TGGAAAGAGC ACCACGGATA TTGTAAAGCT CTTCTGAGAA TTCATCCTCA AAAGATTCGC 1800

CGTAGATTGA CTCGCCACCC ATACCAGTTC CAGTTGGGTC TCCACCTTGG ATCATAAAGT 1860

CCTTGATAAT ACGGTGGAAA ATGACACCAT CATAGTAGCC ATCTTTTGAA AGAGATACAA 1920

AGTTAGCCAC TGTTTTAGGA GCATGTTCAG GGAAAAGCTT GATACGTAAG TCTCCGTGAT 1980

TGGTCTTAAT AGTCGCAAGA GGACCTTCTA CTGTTTCAAT GTCTACTTGT GGAAAATGCA 2040

ATTCTTTTTC TACCATACCA AATACTTCTA AGGCAGCAAA AATGCCATCT TCTTCTAATG 2100

TTTTTGTAAT ATAATCTGCT TTTTCTTTGA TTTTATCATG AGAAATTCCC ATGGCAACGC 2160

TGATTCCAGC ATAATCAAAG AGTTCCAAGT CGTTGAGACC ATCTCCAAAA ACCATGACCT 2220

TCTCTGGTTT CAAGCCAAGG TGTTCCACAA CCTTTTCCAC CCCCGTCGCT TTGGAGCCTG 2280

AAATCGGCAC AATATCAGAC GAATGTTGAT GCCAACGAAC CATGCGAAGT TTGTCTGAGA 2340

GACTGTCAGG CAAGTGCAAG TCATCTCCCT TATCTTCAAA AGTCCACATC TGATAGATAT 2400

CTTCTTTTTC ATGGAAATCG GGATCTACAT CTAAGTCGGG ATAAATTGGA TTGATAGCTT 2460

CACTCATCAT ATCGGTGCGA GTCGACAACT TGGCATCATG ACTCCCAACC AAGCCATACT 2520

CAATTCCTTC TTGCTTAGCC CAAGAGATAT ACTCCTCAAC ATCTGACTTT TCAATCTGAT 2580

GCTGATAAAT GACCTGACCT TTTTTATCTT CGATATAAGC CCCATTCAAA GTTACAAAAA 2640

AGTCAGGCTT GAGATCACGA ATCTCTGGAA CAACACCAAA AATGCCACGT CCAGAGGCGA 2700

TTCCTGTTAA AATTCCTTTT TCACGCAACT GTTTAAAAAC AGTGGGAATT GTAGTTGGAA 2760

TAAACCCTGT CTTTGAATTC CGCAATGTAT CATCAATATC AAAAAAGACA ATCTTGATCT 2820

TCTTTGCCTT GTATCTTAAT TTCGCGTCCA TCTCACTACC TCTTTCAATC TAACTCTTTC 2880

CATTATATCA TAAAGTAGGC AAATCCCCTA TTTTCAAAAA GTTTATCATT TTTATTTTAA 2940

TTTCTTGGAT GAGAAAAGAG ACATATTTAT GAAAAAGCTC CATCGTGCTT TTAATGTGTT 3000

CTCTTGTTTT CAAACTCGTA AAAAGGGAGC CACTGATCCT AACTCGCTCT CTCATTTCAA 3060

AGCTTGTGAA AAAAGACCCG TTGGGGTCTT AATTCGCTTT CTTGTTTTCA AGCTCATGAA 3120

AAAGAGACCC AACTGGGTCT TTTCTTTAAT CTTCGTTTAC GAAAGGCATC AAAGCCATTA 3180

CGCGAGCGCG TTTGATAGCT GTTGTTACTT TACGTTGGTT TTTAGCTGAA GTTCCTGTTA 3240

CACGACGAGG AAGGATTTTC CCACGTTCTG AAACGAAACG GCTAAGAAGC TCAGTATCTT 3300

TGTAATCAAC ATATTCAATT TTGTTTGCTG CGATGTAATC AACTTTTTTA CGGCGTTTGA 3360

ATCCGCCACG ACGTTGTTGA GCCATGTTTT TTCTCCTTTA TAAGTTTAGT TGTCCATTAG 3420

AATGGTAAAT CATCATCTGA AATATCCAAT GGGTTTGTTG CTCCAAATGG ATTTTCATTA 3480

CGTGAAAAGT CTGGTACTGA ATTTGTAGGT GCTGAATAGT TTGCAGTTGG TGCAGAGTAA 3540

GCTCCACCTG TGTGACCCTC ACGCACACTA CGGCTTTCCA ACATTTGGAA ATTCTCAGCC 3600

ACGACCTCTG TCACGTAGAC ACGTTGTCCT TGCTGGTTAT CGTAACTACG AGTCTGGATA 3660

CGACCTGTCA CCCCGATAAG TGAGCCTTTT TTAGCCCAGT TAGCAAGATT TTCAGCCTGT 3720

TGGCGCCACA TAACGACATT GATAAAATCA GCCTCACGTT CACCATTTTG ACTCTTAAAT 3780

GTACGGTTTA CTGCAAGAGT AAAAGTCGCA ACTGCTACAT TTGATGGGGT ATAACGCAAC 3840

TCAGCGTCAC GTGTCATACG CCCTACAAGT ACAACATTGT TAATCATAGT TTACCTTCTT 3900

ACGCGTCAAT TTTGACGATC ATGTGACGAA GAATGTCAGC GTTGATTTTT GAAAGACGGT 3960

CAAACTCTTT AAGAGCTGCA TCGTCATTTG CTTCAACGTT AACGATGTGG TAAAGTCCTT 4020

CACGGAAATC TTGGATTTCG TATGCAAGAC GACGTTTTTC CCAAGTTTTT GATTCAACAA 4080

CAGTTGCACC GTTGTCAGTC AAAATAGAGT CAAAACGTGC TACCAAAGCG TTTTTAGCTT 4140

CTTCTTCAAT GTTTGGACGA ATGATATAAA GAATTTCGTA TTTAGCCATT GATATGTTCC 4200

TCCTTTTGGT CTAATGACCC CAAGACTTTG CAAGGGGTAA GTGAGGTTCG CTCACAATAA 4260

ACTATTATAC TAGAAAAAAT TTTTTTACGC AAGTAAAAAC ACTAGAATTC GAAAAAACGC 4320

CACATGGGCG TTTTCCTGTT CTTATGGTTT GATACGGTGC AACATACGTG GGAATGGAAT 4380

AGCTTCACGG ATATGTTTTG TTCCTGCTGC GAAGGTTACC ATACGTTCGA TACCGATACC 4440

AAATCCTCCG TGTGGAACTG TACCGTATTT ACGAAGGTCA AGGTAGAATT CATATTCTGT 4500

ACGATCCATG CCAAGTTCAT CCATCTTAGC GACAAGGGCA TCGTAATCTT CCTCACGCAT 4560

AGACCCACCG ATAATTTCTC CATAGCCTTC TGGAGCAAGC AAGTCTGCAC AAAGCACGCG 4620

CTCTGGATTT CCAGGAACTG GTTTCATGTA GAAGGCCTTG ATGGCTGCTG GATAGTTCAT 4680

GACAAATGTT GGCACACCAA AGTGGTTTGA AATCCAAGTT TCGTGTGGTG ACCCAAAGTC 4740

ATCACCATGC TCAAGATGCT CGTAGTCAGC ATCTTCATCA TTTTCATGCT CTTGCAAGAG 4800

GTCAATGGCT TGATCGTAAG TGATACGTTT GAATGGCTCT GCAATGTAGC GTTTCAAGAG 4860

TTCTGTATCA CGTTCCAAGG TTTCCAAGGC TTGAGGCGCG CGGTCAAGAA CACCTTGTAG 4920

AAGAGCTTTC ACATAAGCTT CTTGCAAGTC AAGCGACTCA TCATGTGTCA AGTATGAGTA 4980

CTCAGCATCC ATCATCCAGA ACTCAGTCAA GTGACGGCGT GTTTTTGATT TTTCAGCACG 5040

GAAAACTGGA CCAAAGTCAA AGACACGACC AAGAGCCATA GCCCCTGCTT CTAGGTAAAG 5100

CTGACCTGAT TGGCTCAAGT AGGCTGGCGT TCCGAAGTAG TCAGTTTCAA AGAGTTCTGT 5160

AGAATCTTCT GCCGCATTTC CTGAAAGAAT TGGGCTGTCA AACTTCATAA AACCGTTCTT 5220

GTCAAAGAAC TCATAAGTTG CATAGATAAT AGCGTTACGG ATTTGCAACA CAGCTACTTG 5280

CTTACGAGAG CGTAgCCACA AGTGACGGTT ATCCATCAAA AAGTCTGTTC CGTGTTCTTT 5340

TGGTGTGATT GGGTAGTCTT GAGATTCACC GATCACTTCG ATGTCTGTGA TGTCCAACTC 5400

ATAGCCAAAT TTAGAACGTT CGTCCTCTTT GACAATACCT GTCACATAAA CAGACGTTTC 5460

TTGGCTCAAG CGTTTGATAA CATCAAACTT CTCAAGTCCC ACTTCTTCAC CAAATTTTTC 5520

GACAAAGTTT GGTTTAAAAG CCACACCTTG AAAGAAGGCT GTTCCATCAC GCAATTGTAA 5580

GAAAGCGATT TTTCCTTTTC CTGATTTGTT GGCAACCCAA GCGCCAATCG TCACTTCCTG 5640

ACCAACATAG TCTTTTACGT CAATAATCGT TACACGTTTT GTCATTATTT TTCCTTTTCT 5700

TTTTTATTCT TTATGGCAAA CCACCTCTAT ATTGTTCCCA TCCAGGTCAA TCATAAAAGC 5760

AGCATAGTAA ATCGGATGCT CACTTCGATA ACCAGGAGCC CCATTGTCTC GCCCACCTGC 5820

CTCTAAGCCA GCCTCATAAC AAGCCTGAAC TTCTTCCTTA TTTTCTGCTA AAAAAGCAAA 5880

ATGAACAGGA TCTTGTGTTC CCTGAGTCAG CCAAAAATCA CCACCAGGAT GAGGGCTGTT 5940

CGGGGATAGA AAACTAATTA GAGAACTAGT CTTAAAAGCC AATTTATAGT CCAAAGGAGC 6000

GAGAAAACTC CTATAAAATC CTTATGAAAT TTGTAAATCC TTTACCTTAA TCTCAAAATG 6060

ATCAATCATT CTCACTACCC ATAAATGCTT TCAAGCGTTC GACTGCTTCT TTAAGCGTGT 6120

CTAGGTCTGT CGCATAGCTG AGGCGGACAT TTTCTGGTGC TCCAAATCCA GCTCCTGTTA 6180

CCAAGGCCAC TTCGGCTTCT TCTAAGATAA CAGTTGTAAA GTCTGTCACA TCCGTGTAGC 6240

CTTTCATCTC CATGGCCTTT TTGACATTTG GGAAGAGATA GAAGGCCCCT TGCGGTTTGA 6300

CCACTTCAAA TCCTGGTACC TCTGCAAGGA GGGGATAGAT GGTATTAAGA CGTTCCTCAA 6360

AGGCCTGACG CATGCTTTCT ACAGTATCTT GCTCACCTGA TAGAGCCTCA ACTGCTGCAT 6420

ATTGGGCTAC TGCTGACGGA TTCGAAGTTG TTTGACCTGC AATCTTGGAC ATGGCAGCGA 6480

TAATGTCTGC TTCTCCAACG GCATAACCAA TCCGCCAACC AGTCATGGCA TAAGTTTTAG 6540

ACACACCATT GATGACCACT GTTTGCTTGC GAATCGCTTC CGATAGGCTA GAAATCGGTG 6600

TGAACTCATG ACCATTATAA ACCAAGCGGC CATAGATATC GTCTGCTAGG ATGAGAATAT 6660

CATTTTCTAC AGCCCAGTTT CCAATTGCCA AGAGTTCCTC ACGGGTGTAA ATCATACCTG 6720

TGGGATTAGA TGGCGAATTC AGCACCAAAA CCTTGGTCTT GTCAGTGCGA GCTGCTTCTA 6780

ACTGCTCTAC GGTCACCTTA AAGTGATTGT CTTCCTTAGC AGAAACAAAG ACGGGAACGC 6840

CTTCTGCCAT CTTGACCTGA TCTCCATAGC TAACCCAGTA TGGGGTTGGG ATGATGACTT 6900

CATCACCTGG ATTGACCACA GCCATAAAGA AGGTATAGAG AGAATATTTG GCTCCCGCAG 6960

CGACTGTCAC TTGATTTGAC GCTACAGAAT AGCCGTAAAA GCGCTCAAAG TAGCTATTGA 7020

CCGCCGCCTT AAGCTCTGGC AGACCTGAGG TTACTGTATA AAAAGAAGCA CGCCCATCTC 7080

GAATCGATGC AATGGCGGCA TCTTGGATAT TTTTGGGAGT AGTGAAATCT GGCTCACCCA 7140

AGGTTAGAGA CAAAATATCT CTACCCTCAG CCTTCAGTGC TTTGGCACGG GCTCCAGCAG 7200

CCAAAGTCAC ACTTTCTTCC ATTTCTAAAA CACGGTTGGA TAGTTTCATA GGCCCTCCTT 7260

GTTGACCAAT GCTCCTGTTT CAAAATCTAC TAGATAAAAA TCAGATCCTG ACTTAACTTC 7320

CCAGATTGGC TTATCTTGAT AACGGCCAAA GGTTATCTTG TCAATCTCGC CAGCTCCCTT 7380

TTCCTTAGAA ACCGTTTCTG CTTTTTCTTG TGAAACACCC TGATTTAGCT GATAAACGTA 7440

AATCTTATGG TCATCTTTAC CAATCAGGAC AGCAAGCGCT TCTTGCTGTT TGTTACGACC 7500

AAGAACGCTG TAATAAGATT CCAAGCCATT GTATAAATCA ACCTGATCAG CCTGCTCTAA 7560

TCCTGCATAC TGCTGAGCTA ATTTTTCTCC TTCACTTTTA GCTGTTTGAT AGGGTTTCAT 7620

GCTAAGAGAA ACCATATACA GAAAGGAACC ACTGATAACC ACAAACAAAA TCGTCATCCC 7680

TAGACCATAC TGCCACAGTA GATTATTTTT TGCTTTGTTT TGTCTTTTTT TCACTCGTCT 7740

ATTTTACCAT CTATTAAGCT TTATTACAAG TGAATATAAG AATACTCTTC GAAAATCTCT 7800

TCAAACCACG TCAGCTTTAT CTGCAGACCT CAAAGCTGTG CTTTGAGCAA CCAATTCTAT 7860

TTCTCCCTTC AAACAAAACC GATTTTGAAA GTGAAACAGT TCTTACTTTT TCAGTCACAA 7920

ATGATTAGAG TTTGCCGGG 7939

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9897 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CCGCTCTACC GTCAAATAAT TACCATTTTG TTTAATACCG AAATTTTTAT CTACTGAAAA 60

TTCAGTTGGT CTGTTGGTAC GATCGTCGTA TACAGTACCA TTCTCACGAA TAGTATAATT 120

GTAATCAGTA TCACCTTGTT TCCTTAATTT AAGGTAATAA TTACCATCAA TTTGTTTATA 180

ACCTGAATCT TTTCTAGTTG CTTCTCTAAA ACTTACTCCA GCAGGCATCA CATCAGCAAA 240

CATGAGTACT TGTTTGTTCT TTTTTTCAAC AATAACAGAG TCAATATAGG TTGCACCACC 300

GCTGATTTGT AAGTCACGTC CACCAACTTC ACGAGGCCAT TCTAATGGTA CTGGCGCAAA 360

ATCATCGAAT GCCAATGTTA ATTTTGGTTT AGTCCATGTC TTACCATTAT CATCACTATA 420

ACTTGTAGCA ATATTAATTT TATTCAAGAA ATCATGAGTT CCACCGTAAC GAGCGTCAAT 480

GCTTGAAAAT ACCCGACCAT TGCTAAAAGT ATACAGAACT GGAATACGGA AATAGTTAGA 540

ACCTGTTGTA TCATTAGCCG TATAAATTAA ATGTCCAGTA ACAGCGTTTG TTGTCATCTT 600

TTTAACAGTT TCTTCATCCA ATGCACTATT AAAGAATTTG ATATTTTCTA GTGTTCCGTT 660

AAAACCAAAC GCCGTTTTTC CTGCACGTTT CACTCCCCCA AGCATATAGT AATCAATACC 720

TTTAATATCC TTGATGTTTA GGAAATTATC CACTTTCTTT TCTACTACTT TTGTACCATT 780

TGCGTATAAA GAATATGTTT TTTTGACTGA ATCTGCTACT ACTGCAACAG TGTTAGTCAC 840

AGCCTCTTGT TTGTACTTAC CCCAAACTGA AGCAGGTCTG GATACTAGGT TATTTTTATT 900

GGAAGAAGTA TCACGCGCTT CCATCCCCAA CTCACCATTG TCTCTAAGGA ACACATCTAC 960

ATAACTATTT TGTTGACCGG GTTTGGAATT AGATATTCCA AACAGAGCTT GTAAGCCTTT 1020

CTCACTTGAC TGATTGTACT TAATCACTAC AGTAAAGTCA CCGCTAGTAA ATTTATCCTT 1080

TAACTCTTTA GTAACATTTT CTCCGCCCCC TGTTAAAGTA ACATTATTTT TTTCTAAGAC 1140

AGGAGTTTCT TCCGCTGTAG AAGATGGATC CTTAACAGTA GTTTCAACTG TTCGAGGTTG 1200

TACAGTAACT TCCGAAGAGT TATCCGATGT AGGTTGTACT TCCGAAATCG GAGTCGTTGG 1260

TGCAACAGGT TGCACCAACT TTGGTGTTGA TACTTCAGAA GTTTCAGTCT CCTGAGCTGC 1320

AACTGAGTTA GCAACAAATG CTGATAATAC CACTACAGTA CCTAAGGTTA CATATTGTTT 1380

AATATTTTTT TTCATTTTAT TTTTCCTCGT TTAAAACTTT GATAACAAGT TTTTTAACAG 1440

TTTCATCATT GCAATGAATC TTTGGTTGGT GAAGATCTTC TTCAAAAGTC ACCAACATAT 1500

TCCCTGGAAG CAATTCAACA ATTTGATAGT CTTTGCTATC GTAAAAAGCA ATATCCTTCT 1560

CTTCGCTAAA AGGTACACGT GACTGGGCAC GAACTGGGGA AGTTACTGCC ATTTTTTCAG 1620

TATTTTCAAC AACAATATGA ATATCTAAAT ATTTCTTATG AGTTTCAAAA ATATCTCCTG 1680

GAACTCCATC AGCTAGATAA GTCATACAAT TTGCAAAAAC ATTTTCCCCG TCAATATCAA 1740

TTTTTCCATC AACTAAATCT GTCAAATTTG TATTTTCTAA AAAATCACAG ACTTTTGAAA 1800

AATATTTATT GACAGAAGCA TATCGTTTAA AATCAGATTG TTCAGAAATA ATCATATTAT 1860

TTTCTCTTTT CTATTAGTGA CGAACTTCCC AACTTGAATC CGCTTTAATT TCTGTAATAT 1920

CATGAATCGT TGTATATTTA GGTGCAGATA CTTTATTTCC AGTAAGAACA GATACAATAT 1980

AACCTGAAAC TACTGATACA GAGATTGAAA TCAATGAATA TGCCCAGTAG CTAACAGCTG 2040

TTGGAGGAAG GAAGTATTTA ATAAATACCA TGACGATGGT TGATACAATC AGCGCTGCAT 2100

AAGCACCTTG TTTATTTGCT TTTTTAGAAA CAAATCCAAG AATAAATACA CCACCAAGTA 2160

GACCAAGTAC AAGTCCCATG AAACTATTGA ACCATTCGTA TGCAGATTTA ATATCTGAGT 2220

GAGCCATGAC AATGGAAACA CCAATTGAGA ATAAACCTAC TGCTAGAGAT ACGAATTGTG 2280

CAATTTTCGT ACGACGATTG TCTGACATAT TTTTAGAAAT GACATCTTGA ATATCCAATG 2340

TCCATGAAGT TGCAACAGAG TTCAAACCTG TTGAAATAGT TGATTGAGAT GCTGCATAAA 2400

TCGCTGCCAA GATCAAACCT GTGATACCTA CTGGTAACTG GTATGCAATA AAGTACATAA 2460

AGATTTGGTC TTGAGGGATA TTGCTAGCTG CACTATCTGC ATTTTGTACT TGATAGAATA 2520

CGTACAAGCC TGTACCAATC AAGTAAAAGA CTGTTGCAGT TGCAAGTGAC AAAACACCGT 2580

TTGTGAACAA CATCTTATTA AGTTTCTTAA TATTTTGTGT TGTAGTAAAA CGTTGAACCA 2640

AATCTTGAGA TGAAGCATAG GAAGACAAGA TTGTAAAGCC TGAACCCATC ACAATTAAAA 2700

AGATGGAGTT TGAAAGCAAG TTAGGATCGA AAAGTTTTTC ATTTGCAGCA AGGAATTTCC 2760

CGTTTGCTAA TGTTTCTGCT ACTGCACCAA AGCCACCTTT AATATTAGCA ATCAGTACAA 2820

ATAAAGCTAA AACGACACCA CTAATCAGAA TCACACCTTG AATAAAGTCT GTCCATAATA 2880

CGGATTTTAG ACCACCAGTA TAAGAATAAA CAATTGCAAC TACACCCATC AAAATAATCA 2940

AAATATTGAT GTCAATTCCT GTCAATACTG ATAAACCAGC TGATGGGAGG TACATAATGA 3000

TAGACATACG TCCCAATTGA TAAATAATAA ACAAGAGTGC TGAAATAATA CGAAGTGCTT 3060

TAGAATTAAA ACGTTTATCC AAGTAATCAT ATGCCGTATC GATGTCTATC CGTGCAAAGA 3120

TAGGTAAGAT AAAACGAATT GTCAGTGGAA TAGCTACTAC CATCCCTAAT TGAGCAAACC 3180

ATAAAATCCA GCTACCTGCA TAAGAGCTAC CAGCGAGTCC CAAGAAGGAA ATCGGACTGA 3240

GCATTGTGGC AAAAATGGAT ACCGAAGTAA CATACCAAGG AACCGAACCA TCTCCTTTAA 3300

AGAACTCTTT TCCTTTCATC TCTTTTTTAG AGAAATAGAT ACCTGCAACC AACACCGCAA 3360

GTAAATAAAC AATCAAGATA ATTAAGTCAA TTATTGTAAA TCCTGTTGTG CCCATAACAT 3420

ATCTCCATAT TGATTTTATT TATTATAAAA ATTCTTTTCG TGCTTGTTGA ATAAGTTCTG 3480

CTGCTTGTTT TGCAACTTCC AAGTCACCTT CTGCCAATGC TTCTAAAGGT TGACGAACAG 3540

AACCTAAATC AAGTTTTTCA TTTAGACGCA AAACTTCTTT TGCTACAGCA TACATATTTG 3600

CCTTACCTGA TATCATCTTA TAGATAACTT CATTGATAGC ATATTGAAGT TTTTTAGCTG 3660

TATCTAAATC TCGTTCTTGA ATCAAACTTT CCAATTTCAA GAACAAATCT GGCATAACGC 3720

CATAAGTACC ACCAATACCA GCTTCTGCTC CCATCAAGCG ACCACCAAGA TATTGTTCAT 3780

CTGGACCATT GAATACAATG TAATCTTCTC CACCTGCAGC TACAAACATT TGAATATCTT 3840

GTACAGGCAT AGAAGAATTT TTAACTCCAA TCACACGAGG ATTTTGACGC ATTGTTGCAT 3900

ACAAACTACC AGTCAACGCA ACCCCTGCCA ATTGTGGAAT ATTATAGATA ATAAAATCTG 3960

TATTTGACGC AGCTTCACTC ATTGCATTCC AATATGCTGC GATTGAATAC TCTGGCAATT 4020

TGAAATAAAT AGGTGGGATA GCTGCAATAG CATCGACTCC AACACTTTCT GAATGTTTTG 4080

CCAATTCGAT ACTATCTTTC GTGTTATTAC ATGCAATATG GTTGATAACT GTTAATTTAC 4140

CTTTAGCAAC TTCCATAACA GCTTCAATAA TTTGTTTACG ATCTTCTACA CTTTGGTAAA 4200

TACATTCACC TGAAGAACCA TTTACATAGA TACCTTTTAC ACCTTTGTCA ATGAAATATT 4260

GTACCAGAGA TTTTACACGA TCTTGGCTAA TTTCACCATT TTCATCATAG CAAGCATAAA 4320

ATGCAGGGAT AACGCCTTTG TATTTAGTTA AATCTTTCAT CAGATTTCTC CTTTATATTG 4380

TTTTTTATTT GATGACATTA ATAAATCGCT GAGCAATTTC TTTTGGACGT GTAATCGCTC 4440

CACCAATGAC TACACTGGTA ACACCTAAAC TATAAGCTTT TTTTAATTGT TCTGGATAAT 4500

GAATTTTTCt TCGGCAATTA CCGGAATATT AAAATCAGCC AATTTTTTCA TTAGTTCAAA 4560

ATCAGGCTCA TCTGATTGTA CACTTGTACT TGTGTAACCT GATAATGTTG TACCAACAAA 4620

ATCAACGCCT GATTTAAATG CATAGAGACC TTCATCTAAA TTACTTACAT CCGCCATCAG 4680

CAATTGATTC GGATATTTTT CTTTTATTTT TTTGATAAAT TCACTGACAA CTAAGCCATC 4740

ATATCTTGGT CTTAAAGTTG CATCAAATGC AATGACTGTT GTTCCGCATT CTACAAGTTC 4800

ATCTACTTCT TTCATCGTAG CAGTAATATA TGGTTCTTGA GGTGGATAAT CCCTTTTGAT 4860

AATTCCAATT ATTGGTAAAT CTACTACTTT CTGAATTGCT TTAATATCAC GCACAGAATT 4920

TGCGCGAATG CCCACTGCTC CTGCCTCTAA AGCTGCTTTA GCCATAAAAG GCATCAAGCT 4980

AAATTCTTCA TTATAAAGGG CTTCACCAGG TAAAGCTTGA CAAGAAACAA TGACTCCACC 5040

TTGAACTTGG CTTATAAATT TTTCTTTAGT CCAAATTTGG CTCATTTTAT TATTCCTCCT 5100

TATGGATAAT AGTTTGATTG TAATAATATT GTCTCTCTGG ACTTTCCAGA TAATTAGAGA 5160

ATAAGCAGTC TGTAATTAAA AGTATTGGAA ACTGAGGTGA TATGCGATTG CCATACGAGA 5220

GATGATCGGT CGAAGCTAAT AACAATAGTT CATCAAAGAA ACAATCTTCT TCGTCAAATT 5280

TTCTTGTAGT CATTAAAACT GTTTTAGCGC CTTTATCTGC AGCTTTTTGT AGACCTTCTA 5340

GTACAATATC AGTTTGACCT GAAATGGATG CTCCAATGAC AAGGCAATTT TCATTAAGTA 5400

GTAAGCTACT CCACAAAATC ATATCCTCGT CTGATAATAC TTCACCAATC ACTCCGAGAC 5460

GCATAAATCT CATCTTCATT TCTTGTAAAG CAAGAACAGA ACTTCCTTTA CCGTAGAGAT 5520

ATACACGCTC AGCAGTTTCT ATCATCTCAG CAATACGCTC AAGTTGAACT TCATCAAGAA 5580

CCGTGTAAGT TTTTCTCAAC ATTTCCTCAT AGTCGGATAA AACTTTTTCT GTTGCCTCTG 5640

TATATAATGC CAACTTTTCT TTCTCATGAA TCATCTCTTG GTATTTGAAA ATGAATTGTC 5700

TAAAACCTTT AAAACCACAT TTTTTCGCAA ATCGAGTCAA TGTTGCTTTG GATACATTAA 5760

GGTATTCGCA CAATGCTTTA GATGAATAAT CATTCAGAGG TTGCTGTTTT AAGAAGAATT 5820

TAGCAATGTC TTTTTCAGCA TATGCCATAT TTGGTAAGTT AGCTTCTATC ATTGGAATTA 5880

GTTCTTTTTG CAGTAACATA TGAGCTCCTT AGTTGAAGTA AACGTTTACA TTCTTTATTT 5940

TAACACTTTT TTTTTTTTTC AATATTTTTC ATAAATTAGA AACTAGTTTC CAATTTCTTT 6000

CGTTTCATAA CAGAACAACA AACATAAAAA TATAATAGTT TTTATTCTTT TTATCGTAAT 6060

TATATGTATT GTAAGAACGT TTATCACTAA TAATATGTTC ATATTAAAAT ATTTTAGTAA 6120

TATTTTATTT TGGTTTTATT ATTTCTTTTC GGAATTTCTA TATAATATTT TATTTCTAAA 6180

AAAATTGAAA AAATATTTCT AGTTTCTTTA TTTTATATAG GTAATATATT TTATTTCTAA 6240

ATTAAAAGAG AATCCCATAA AAACTACAGA TTTATGAGAT AAATCAGGTC ACCTATTTTA 6300

AAAAAGCAGC AAACTATAAA CTAAAAAGTT CCACACCAAA TGTAACCCCA TACTTCCCCA 6360

TAAGTCAGAT TTATAGCGCA CCATACCTAA AAACATTCCA AGTGAAACGT ACAGACACCA 6420

AGCTAGAATG GTTCCTGGAT GATGTACTAA GGCAAATAAA ACACTTGTCA AAGCAACTCG 6480

AATATCTAAT TTTCTAACCA AGTTCCATAA AATTTCACGA TACAGAAATT CTTCAACCAT 6540

ACTCGCATTG ATTAAGAACA ATAAAAATGA AAACCAAGGA ACTTGATGTT GAAGGCCAAT 6600

TAAATTTGTT TGATTCGTGC TTCCTTGAGC ATGAATCAGG CTAAAACATA GACTTATAAT 6660

CAGTAGACTA GCTAGTCCAA TACCAAGGCA TTTCATCCTA GTTTTCATAT TGACCTTGAC 6720

CACTTGTTTT CGTTGACCAT ACATCCATAA AAAAGAAAAA AGAGACGCAC CATAGAGAAC 6780

CTGTAGTATA GTTAACTCAC CGATACAAAG AAATTTCAAT AAGTATAGAG ATACCAATAG 6840

GACATTTACT TGTTGGAATA TATAAACTGG AATTATTCTT TTCATAGTTA CCTCCGAAAT 6900

AAATCTTCAT AATCTAAATC TAATATCTGC ACAATCCTTT CTACCCATGG ACTTTGAGGC 6960

ATTCGTTGTT CCATCTTGTA GTGGCGAATC TTTTGATATA AACGATTCAA TTCACTTGGA 7020

TAGTGAAACT CTCCCGCAAA CATTTTTCTG GTTAACTCAA TCCAGCTGAT ATTTCTTTCA 7080

GCCAAAATAA TGGACAAGTT CTCCCAAAAT CGTTCAGCCA TATTrCTTCT CCTTTAGTTA 7140

GATAAATAAT GTGTTTGyGC CATGTAAATC AATTGTTTCG TATCTCTTGG CAATAGAGCT 7200

CTAGCCTCTT CCAAATTCAG ACTTGGATAA ACCCGCTTAT TTGAAACCAC AAAAGGAAGT 7260

CCGATGGTTA GTTCAGGATT TTTTAAAATT ATCTCAACGA AATCCGTTAA TCTTAGATTG 7320

TCACGGTTCT TAAATCGTAA TAAATTGGGA GATAAAAACT CAAAACAATC TGAAGAATAG 7380

CTCATCATCT CAATTAATTT GTCCTTTGTC ATTTCAGAAA CTGAATGACA AGATACCTCA 7440

ATGCCATAGT TTTGGAAGAA GTCTAAAAGA AGTTGATTTC TTTGGCTATT TTTACTTAGA 7500

TAGAGATCAA TCATGGGAGA CCTCCAACAA ATTTGCTTCC ATTTGATATT CTGAGACGAT 7560

TAAGGAATCT AACAACTTTG AGAAGTTAAT CGATTTCTTG TCTTCATCAT AAGCTTTTAC 7620

AGTTACTTGG GTTGTAAGTA TCCCCTCTTT TCCCTCGGCT CGATAGTCTT GTCAATATAA 7680

AACAAAAACA AGATTCTGAT TATCATCTAC AAAGGCATTA ACTCCGTTCT TTATATCCTG 7740

ACTTTCAAGG AATTCCATAA CGTTTTGAAG ATAGGATTCA TAAAATAGTG GGTAATTATG 7800

TTTTTTATGG TAATCATCTA AAAATGTTAC CTCAAACTCA CATGGATAAT TGGGCATCAA 7860

AAATATTTGT TCATCCAGCT GTTTGATTTC TGCATCATGT AATTCTGTTT CTAATTCATC 7920

ACAATCTAGT ATTGATTCTT TATTTAATGC TTTTATCTTT TTCCTCTATT TCTTTTAATT 7980

TCTTTGCGAT TGCGGCAATC ACAGGAACGG TTACACTATT ACCAACTTGT TTATAGAGCT 8040

GACTATTAAT AGAGACTTTT CTAGCAGCTT CAAAAGCCTA ATCAGGAAAG CCATGCAATC 8100

GAAAACACTC TTTAGGAGTG ATTCGTCGTA TTCTCAAACG GTAAAATTGT CCATCTATTA 8160

AAACACCAGC TACTTGGTAA ACTTGTTTAT CTTCTCCTTC ATAGCTAGCC ACTACTACTC 8220

CCATTTGACC ACTAGTTGTT AACGTATTAG CTATACCTTT TCCAACTCTA CCACGACGAT 8280

ACTGAGAACT TGGTCTTTCT AAATTGATTG AATCCCCAAT CTCTGCTTGA GCATATCCTT 8340

TTTTCGTTGC TTCCCGTACT TTTAGAAATT GGATTGGTTC TGGAATTAGT ATTTTGGGGA 8400

TTTTATCTCC TCCTTGCATC GTAGTCAGTG TTGGAGATAA GCCCTCACTT CCATAGACAC 8460

GACCTGTCTC CTTAAAGCTA GTCGGTAAAT CTCCAACAAC GACAATGCCA TAACGATCCT 8520

GAGTATTTAA AGTAAACATC GGCTCTTGAT TTTCCTTAAA GCGTCTCCCA TTTTGTCTCT 8580

TGTCTAATCT ATCTGGTGTC ATACAAGGAA TCGCAACTTT AAATCCTTCT CCTTTACCAC 8640

GAACTAAGGT TGGCGCAAGA CCTTCTGAAT AATAGACTTT ACCGCTCATT CCACTTCTTG 8700

ATGGATTCAA ATTTCCTAGT GCTTTCAAAG TCTCAGAGTT AGTTGCTTGA CCTTCTCGTC 8760

TGAAAGGAAA TAAGAGTCTG GTACCTTTCT TTCTAGAATG TCCGATAATA AACACCCTCT 8820

CTCTGTTTTT GGGAACGCCA AAATCCTTAC TGTTAAGCAC CTGCCACTCA ACATCAAACC 8880

CCAACTCATC AAGTGTGGTA AGTATTGTGG TGAACGTCCG TCCCTTATCG TGATTGAGTA 8940

GGCCTTTAAC ATTTTCAAGA AAAAGAAAAC GTGGTTGGAT TTGTTTGGCC GCCCGAGCAA 9000

TTTCAAAGAA CAAAGTTCCT CTAGTATCTT CAAATCCCAA TCGTCTTCCT GCGATTGAAA 9060

ATGCTTGACA AGGGAATCCC CCACAGATGA CATCGACTTT CCCTCTAAGT TTTTTAAATT 9120

CGTCATCTGA AACATCTCGT ATGTCATGAA ATTCTATTTC TCCTTCCGTT TGAAAAATGG 9180

ACTTATAAGA TTTCCTAGCA AATTTATCAA TCTCACAAAA TCCCAAGCAC TCATGCCCTT 9240

GAGCTTCCAT TCCCATCCTA AAGCCTCCTA TCCCAGCAAA TAAATCTAAA ACCCAAATCA 9300

TTCATACCTC TCTCAACTAG ATGTAACTTA CAAAACCCCT GACCTCATGA GCCACTTTCT 9360

TCCTCCTCAT GAGGTCAGTT TTACTTTCTG CTGTTCCAGT ATCGTTTTTC CTCGCTAGAT 9420

TTCCTCAAAA GGGCAGACTC CTCCCTTGGT TCGTCACACG ATTTTTTCAT CTCGACTGTT 9480

CTTTAATGCA TCATTAACGA CGCTTTTCTT CTAGGTGGTT CATAAGGAAC AGGAAGATTC 9540

AGGTTGACTT TTCTAATCCT AGAATAAAGT GCTGAAAACA ATTCGGAATA GGCATAGAGA 9600

CTAGACAATT TGAGGAGCTG CTTGCGTCCT GTTCGAACAC ATTTTCCTAC CACGTGAAGA 9660

AAAAGATGGC GGAAGCGTTT GATTGTTAAA GTTTGGAAGT CACCTCCAGC TAGATGTTTG 9720

AGAAAAAGAT AGAGATTGTA GGCGATACAG CTCATCATCA TACGAACTCG TTTTTGATTA 9780

AGGTTGAACT ATCCGTTTTA TCGCCAAAAA ATCCCTCCTT CATCTCCTTG ATGAAATTCT 9840

CGGCTTGACC ACGTCCACGA TAAAGCTGAA ACTGGTCTTG GCTTGTTCCG GTACCGA 9897

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8148 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CCGTGGAACA AGCCAAGACC AGTTTCAGCT TTATCGTGGA CGTGGTCAAG CCGAGAATTT 60

CATCAAGGAG ATGAAGGAGG GATTTTTTGG CGATAAAACG GATAGTTCAA CCTTAATCAA 120

AAACGAAGTT CGTATGATGA TGAGCTGTAT CGCCTACAAT CTCTATCTTT TTCTCAAACA 180

TCTAGCTGGA GGTGACTTCC AAACTTTAAC AATCAAACGC TTCCGCCATC TTTTTCTTCA 240

CGTGGTAGGA AAATGTGTTC GAACAGGACG CAAGCAGCTC CTCAAATTGT CTAGTCTCTA 300

TGCCTATTCC GAATTGTTTT CAGCACTTTA TTCTAGGATT AGAAAAGTCA ACCTGAATCT 360

TCCTGTTCCT TATGAACCAC CTAGAAGAAA AGCGTCGTTA ATGATGCATT AAAGAACAGT 420

CGAGATGAAA AAATCGTGTG ACGAACCAAG GGAGGAGTCT GCCCTTTTGA GGAAATCTAG 480

CGAGGAAAAA CGATACTGGA ACAGCAGAAA GTAAAACTGA CCTCATGAGG AGGAAGAAAG 540

TGGCTCATGA GGTCAGGGGT TTTGTAAGTT ACATCTAGTT GAGAGAGGTA TGAATGATTT 600

GGGTAAATAC AATGAGCTTG AAAGAAGTAG CAAACTCACC AAGCGCCAAT TCTTTGAGAA 660

TCAGATGCTG GATTATACCA TCATTGCGCA TGAGAGTTTT GAAATCATCC GTCATTCTGT 720

CTACCAGACA GATGATCGTG AAGTGGAAAA TGCTCTGGCT TTTGAAGTGA AAAATGATGA 780

AACAGACAAG CTGATTCTGT TATTAAGCGA GGATATTGGT GTAGGTGAAA AATTGTGCCT 840

CGTTGACGGA ACAAAAATGC GTGGAAAATG TTTAGTATAT GATAAAATAA ATGAGAGAAT 900

GATTCGCTTG CAGTGCTAGA AATAGGCATT TTGAATAGTG AATATGTTAT AATAAGTATT 960

AGTAGGAGGT GTTTTAGATT GGAGAAGAAA CTGACCATAA AAGACATTGC GGAAATGGCT 1020

CAGACCTCGA AAACAACCGT GTCATTTTAC CTAAACGGGA AATATGAAAA AATGTCCCAA 1080

GAGACACGTG AAAAGATTGA AAAAGTTATT CATGAAACAA ATTACAAACC GAGCATTGTT 1140

GCGCGTAGCT TAAACTCCAA ACGAACAAAA TTAATCGGTG TTTTGATTGG TGATATTACC 1200

AACAGTTTCT CAAACCAAAT TGTTAAGGGA ATTGAGGATA TCGCCAGCCA GAATGGCTAC 1260

CAGGTAATGA TAGGAAATAG TAATTACAGC CAAGAGAGTG AGGACCGGTA TATTGAAAGC 1320

ATGCTTCTCT TGGGAGTAGA CGGCTTTATT ATTCAGCCGA CCTCTAATTT CCGAAAATAT 1380

TCTCGTATCA TCGATGAGAA AAAGAAGAAA ATGGTCTTTT TTGATAGTCA GCTCTATGAA 1440

CACCGGACTA GCTGGGTTAA AACCAATAAC TATGATGCCG TTTATGACAT GACCCAGTCC 1500

TGTATCGAAA AAGGTTATGA ACATTTTCTC TTGATTACAG CGGATACGAG TCGTTTGAGT 1560

ACTCGGATTG AGCGGGCAAG TGGTTTTGTG GATGCTTTAA CAGATGCTAA TATGCGTCAC 1620

GCCAGTCTAA CCATTGAAGA TAAGCATACG AATTTGGAAC AAATTAAGGA ATTTTTACAA 1680

AAAGAAATCG ATCCCGATGA AAAAACTCTG GTATTTATCC CTAACTGTTG GGCCCTACCT 1740

CTAGTCTTTA CCGTTATCAA AGAGTTGAAT TATAACTTGC CACAAGTTGG GTTGATTGGT 1800

TTTGACAATA CGGAGTGGAC TTGCTTTTCT TCTCCAAGTG TTTCGACGCT GGTTCAGCCC 1860

TCCTTTGAGG AAGGACAACA GGCTACAAAG ATTTTGATTG ACCAGATTGA AGGTCGCAAT 1920

CAAGAAGAAA GGCAACAAGT CTTGGATTGT AGTGTGAATT GGAAAGAGTC GACTTTCTAA 1980

AATGAAGGAA AATGACTTGC AATCTCTGTT AAGAAATAAA ATAATCCCAC CTAGAACAAG 2040

CTAGGTGGGA TTATTTGCCT ATGAAATGAG AAATTATGGG AGCAAGCTCC TAAATCAACT 2100

GTTTTTGATC TACTTCTTTA ACTACTTGAT AAAAGTTATA GAAGTAGGCC AAACTTGAAA 2160

TGATGGTTAC GACTAGGAAT ATTGAAAATT TCCATTGGAC AGGGTTGGTT AAAAGTTGTG 2220

GAAAGGATAT GAGGAGAAAG AAGAGGGCTG CGTTGAGGAC AGGTATCCGT TTTGATTGTA 2280

TTTTCTCAAG TCCTTTATTG AGCGCAGGAA GAAAGAGGAG TAGGAGTAGT AAAACTGTAT 2340

GAGAAATAGC TCCTGAAGTA AGGGCGAAGA AAAGGAAAAT ACTGATAAAA ACATGAATGA 2400

TCAGTAGTCT AGCTAGTGAT TTCATAAGGC ACCTCCTAAT CCTGGTCTTT TTTAGCTCTT 2460

GCAATACGAA GTGAGTCGAC AATATGTATC ATCACTCCGA AAAAGAAAGC TCCCAGTATA 2520

GTTTTAAAAA TATGTTTTGT ATTTAGAAGA GAACTGATAA AATTTGGATT TTCACTTGTT 2580

AGGGTATCAA TGAGTGGAAT TATAAAAAAT ATCACTGTTC CATAAATCGA ACCTGCTTTC 2640

AGACCAGGAT AACGTAACTG TTTCTTTTCT TTTTTCATGA GTTTCCTCCT AATCCTCATC 2700

TTGATTTTTC TTAGTTTTTG CAATGCGACG GGAGATGAGG AACTGTATGC TCGCTCCGAA 2760

GAAAATAGAA CCGAGAATAC TTGATACACC ATTTCTTATA GTGAGAAGAG AATGAAAATA 2820

GTCCTGACCT TCATCTATGA GTATCCTGAG AAGAGGAGTT ATAAAAAACA TCCATAGACC 2880

AAAGAACAAA CCTGCTTTCA GACCTGGGTA GTGTAGTTGC TTGCTTTCTT TCTCATTCAG 2940

CATATCTGGT TCAATGACTG TGATGCCTGT TTTTTTCATT TGGTAGGTGA CATAGCCAGA 3000

AGCGATGAGG GCAATCACTA AAATCAGAGG AGGATAGATT AGAGCCACTT CTTGAGGGTA 3060

TTTATAGGCC AGAAGGAGTG GAATAAGATT TCCGAAAATC ATCAGATAAA AGAGGATGAT 3120

AAAGACTTGG TTCCCAATAC TATCGGCCTC ACGCCGTTTG TATTCGTCAA GGGGACCAGA 3180

AATACCGTAT GTGCGTTTGA TCAGTTTTTC AGTGAAGGTT TCTTTTTTCA TGAGTTTGCT 3240

CCTTTTTTAA AAATCTTCCT CCCAAAAGAG ACTGTTGAGG TCAGTTTGGA GGCTGCGGGC 3300

GAGATTGAGA CAGAGTTCCA AGGTTGGATT GTACTTGTCG TTTTCAATCA TATTGATAGT 3360

CTGTCTCGAG ACACCGATAT CCTTGGCGAG TTCGAGCTGG GAAATACCCA ATTCCTTGCG 3420

AAATTCTTTC ACACGATTCA TCTGTTCTCC TTTCTGATTT ATGTCGTATA TATTTGACTA 3480

TATTATAGTC TTTTAAACAT AAAGTGTCAA GTATTTTTGA CATATTTTTT GAAGAAATAG 3540

TAGTCTCCTT GTCCTATTTG TCTGACAAGT GCAAGCTGGT CGGATTTGTG GTAAAATAGA 3600

TAAGATATGA CAAAAGAATT TCATCATGTA ACGGTCTTAC TCCACGAAAC GATTGATATG 3660

CTTGACGTAA AGCCTGATGG TATCTACGTT GATGCGACTT TGGGCGGAGC AGGACATAGC 3720

GAGTATTTAT TAAGTAAATT AAGTGAAAAA GGCCATCTCT ATGCCTTTGA CCAGGATCAG 3780

AATGCCATTG ACAATGCGCA AAAACGCTTG GCACCTTACA TTGAGAAGGG AATGGTGACC 3840

TTTATCAAGG ACAACTTCCG TCATTTACAG GCATGTTTGC GCGAAGCTGG TGTTCAGGAA 3900

ATTGATGGAA TTTGTTATGA CTTGGGAGTG TCTAGTCCTC AATTAGACCA GCGTGAGCGT 3960

GGTTTTTCTT ATAAAAAGGA TGCGCCACTG GACATGCGGA TGAATCAGGA TGCTAGCCTG 4020

ACAGCCTATG AAGTGGTGAA CAATTATGAC TATCATGACT TGGTTCGTAT TTTCTTCAAG 4080

TATGGAGAGG ACAAATTCTC TAAACAGATT GCGCGTAAGA TTGAGCAAGC GCGTGAAGTG 4140

AAGCCGATTG AGACAACGAC TGAGTTAGCA GAGATTATCA AGTTGGTCAA ACCTGCCAAG 4200

GAACTCAAGA AGAAGGGGCA TCCTGCTAAG CAGATTTTCC AGGCTATTCG AATTGAAGTC 4260

AATGATGAAC TGGGAGCGGC AGATGAGTCC ATCCAGCAGG CTATGGATAT GTTGGCTCTG 4320

GATGGTAGAA TTTCAGTGAT TACCTTTCAT TCCTTAGAAG ACCGCTTGAC CAAGCAATTG 4380

TTCAAGGAAG CTTCAACAGT TGAAGTTCCA AAAGGCTTGC CTTTCATCCC AGATGATCTC 4440

AAGCCCAAGA TGGAATTGGT GTCCCGTAAG CCAATCTTGC CAAGTGCGGA AGAGTTAGAA 4500

GCCAATAACC GCTCGCACTC AGCCAAGTTG CGCGTGGTCA GAAAAATTCA CAAGTAAGAG 4560

GGAAAAAGAT GGCAGAAAAA ATGGAAAAAA CAGGTCAAAT ACTACAGATG CAACTTAAAC 4620

GGTTTTCGCG TGTGGAAAAA GCTTTTTACT TTTCCATTGC TGTAACCACT CTTATTGTAG 4680

CCATTAGTAT TATTTTTATG CAGACCAAGC TCTTGCAAGT GCAGAATGAT TTGACAAAAA 4740

TCAATGCGCA GATAGAGGAA AAGAAGACCG AATTGGACGA TGCCAAGCAA GAGGTCAATG 4800

AACTATTACG TGCAGAACGT TTGAAAGAAA TTGCCAATTC ACACGATTTG CAATTAAACA 4860

ATGAAAATAT TAGAATAGCG GAGTAAGATA TGAAGTGGAC AAAAAGAGTA ATCCGTTATG 4920

CGACCAAAAA TCGGAAATCG CCGGCTGAAA ACAGACGCAG AGTTGGAAAA AGTCTGAGTT 4980

TATTATCTGT CTTTGTTTTT GCCATTTTTT TAGTCAATTT TGCGGTCATT ATTGGGACAG 5040

GCACTCGCTT TGGAACAGAT TTAGCGAAGG AAGCTAAGAA GGTTCATCAA ACCACCCGTA 5100

CAGTTCCTGC CAAACGTGGG ACTATTTATG ACCGAAATGG AGTCCCGATT GCTGAGGATG 5160

CAACCTCCTA TAATGTCTAT GCGGTCATTG ATGAGAACTA TAAGTCAGCA ACGGGTAAGA 5220

TTCTTTACGT AGAAAAAACA CAATTTAACA AGGTTGCAGA GGTCTTTCAT AAGTATCTGG 5280

ACATGGAAGA ATCCTATGTA AGAGAGCAAC TCTCGCAACC TAATCTCAAG CAAGTTTCCT 5340

TTGGAGCAAA GGGAAATGGG ATTACCTATG CCAATATGAT GTCTATCAAA AAAGAATTGG 5400

AAGCTGCAGA GGTCAAGGGG ATTGATTTTA CAACCAGTCC CAATCGTAGT TACCCAAACG 5460

GACAATTTGC TTCTAGTTTT ATCGGTCTAG CTCAGCTCCA TGAAAATGAA GATGGAAGCA 5520

AGAGCTTGCT GGGAACCTCT GGAATGGAGA GTTCCTTGAA CAGTATTCTT GCAGGGACAG 5580

ACGGCATTAT TACCTATGAA AAGGATCGTC TGGGTAATAT TGTACCCGGA ACAGAACAAG 5640

TTTCCCAACG AACGATGGAC GGTAAGGATG TTTATACAAC CATTTCCAGC CCCCTCCAGT 5700

CCTTTATGGA AACCCAGATG GATGCTTTTC AAGAGAAGGT AAAAGGAAAG TACATGACAG 5760

CGACTTTGGT CAGTGCTAAA ACAGGGGAAA TTCTGGCAAC AACGCAACGA CCGACCTTTG 5820

ATGCAGATAC AAAAGAAGGC ATTACAGAGG ACTTTGTTTG GCGTGATATC CTTTACCAAA 5880

GTAACTATGA GCCAGGTTCC ACTATGAAAG TGATGATGTT GGCTGCTGCT ATTGATAATA 5940

ATACCTTTCC AGGAGGAGAA GTCTTTAATA GTAGTGAGTT AAAAATTGCA GATGCCACGA 6000

TTCGAGATTG GGACGTTAAT GAAGGATTGA CTGGTGGCAG AACGATGACT TTTTCTCAAG 6060

GTTTTGCACA CTCAAGTAAC GTTGGGATGA CCCTCCTTGA GCAAAAGATG GGAGATGCTA 6120

CCTGGCTTGA TTATCTTAAT CGTTTTAAAT TTGGAGTTCC GACCCGTTTC GGTTTGACGG 6180

ATGAGTATGC TGGTCAGCTT CCTGCGGATA ATATTGTCAA CATTGCGCAA AGCTCATTTG 6240

GACAAGGGAT TTCAGTGACC CAGACGCAAA TGATTCGTGC CTTTACAGCT ATTGCTAATG 6300

ACGGTGTCAT GCTGGAGCCT AAATTTATTA GTGCCATTTA TGATCCAAAT GATCAAACTG 6360

CTCGGAAATC TCAAAAAGAA ATTGTGGGAA ATCCTGTTTC TAAAGATGCA GCTAGTCTAA 6420

CTCGGACTAA CATGGTTTTG GTAGGGACGG ATCCGGTTTA TGGAACCATG TATAACCACA 6480

GCACAGGCAA GCCAACTGTA ACTGTTCCTG GGCAAAATGT AGCCCTCAAG TCTGGTACGG 6540

CTCAGATTGC TGACGAGAAA AATGGTGGTT ATCTAGTCGG GTTAACCGAC TATATTTTCT 6600

CGGCTGTATC GATGAGTCCG GCTGAAAATC CTGATTTTAT CTTGTATGTG ACGGTCCAAC 6660

AACCTGAACA TTATTCAGGT ATTCAGTTGG GAGAATTTGC CAATCCTATC TTGGAGCGGG 6720

CTTCAGCTAT GAAAGACTCT CTCAATCTTC AAACAACAGC TAAGGCTTTA GAGCAAGTAA 6780

GTCAACAAAG TCCTTATCCT ATGCCTAGTG TCAAGGATAT TTCACCTGGT GATTTAGCAG 6840

AAGAATTGCG TCGCAATCTT GTACAACCCA TCGTTGTGGG AACAGGAACG AAGATTAAAA 6900

ACAGTTCTGC TGAAGAAGGG AAGAATCTTG CCCCGAACCA GCAAGTCCTT ATCTTATCTG 6960

ATAAAGCAGA GGAGGTTCCA GATATGTATG GTTGGACAAA GGAGACTGCT GAGACCCTTG 7020

CTAAGTGGCT CAATATAGAA CTTGAATTTC AAGGTTCGGG CTCTACTGTG CAGAAGCAAG 7080

ATGTTCGTGC TAACACAGCT ATCAAGGACA TTAAAAAAAT TACATTAACT TTAGGAGACT 7140

AATATGTTTA TTTCCATCAG TGCTGGAATT GTGACATTTT TACTAACTTT AGTAGAAATT 7200

CCGGCCTTTA TCCAATTTTA TAGAAAGGCG CAAATTACAG GCCAGCAGAT GCATGAGGAT 7260

GTCAAACAGC ATCAGGCAAA AGCTGGGACT CCTACAATGG GAGGTTTGGT TTTCTTGATT 7320

ACTTCTGTTT TGGTTGCTTT CTTTTTCGCC CTATTTAGTA GCCAATTCAG CAATAATGTG 7380

GGAATGATTT TGTTCATCTT GGTCTTGTAT GGCTTGGTCG GATTTTTAGA TGACTTTCTC 7440

AAGGTCTTTC GTAAAATCAA TGAGGGGCTT AATCCTAAGC AAAAATTAGC TCTTCAGCTT 7500

CTAGGTGGAG TTATCTTCTA TCTTTTCTAT GAGCGCGGTG GCGATATCCT GTCTGTCTTT 7560

GGTTATCCAG TTCATTTGGG ATTTTTCTAT ATTTTCTTCG CTCTTTTCTG GCTAGTCGGT 7620

TTTTCAAACG CAGTAAACTT GACAGACGGT GTTGACGGTT TAGCTAGTAT TTCCGTTGTG 7680

ATTAGTTTGT CTGCCTATGG AGTTATTGCC TATGTGCAAG GTCAGATGGA TATTCTTCTA 7740

GTGATTCTTG CCATGATTGG TGGTTTGCTC GGTTTCTTCA TCTTTAACCA TAAGCCTGCC 7800

AAGGTCTTTA TGGGTGATGT GGGAAGTTTG GCCCTAGGTG GGATGCTGGC AGCTATCTCT 7860

ATGGCTCTCC ACCAAGAATG GACTCTCTTG ATTATCGGAA TTGTGTATGT TTTTGAAACA 7920

ACTTCTGTTA TGATGCAAGT CAGTTATTTC AAACTGACAG GTGGTAAACG TATTTTCCGT 7980

ATGACGCCTG TACATCACCA TTTTGAGCTT GGGGGATTGT CTGGTAAAGG AAATCCTTGG 8040

AGCGAGTGGA AGGTTGACTT CTTCTTTTGG GGAGTGGGAC TTCTAGCAAG TCTCCTGACC 8100

CTAGCAATTT TATATTTGAT GTAAGAATGG CACCCTGATG TTTCAGGG 8148

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9909 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

TACTCCACCC TTAATATCCG TTCCTGTAAA TACTTTACCG CTTTTAAGTT CATAGAATTG 60

AACTTTTAAA TGCTTGTCTT CAAGCATCTT TTCCATCCAA TTTTTAGGAG TTTGACCAGC 120

TTTAAATAAA AACCTTGCTG GGGTGATTAG TATAGATTTA TCTGCGATTT TATAAGCTTC 180

ATCAATAAAA TAGTGATATA TCGGCTCATC TCTGGCTTCT CCTGTTTCCT GATACGGAGG 240

ATTTCCTATC ACGACATCAA ATTTCATTTC ACTTTCCTCG CTAGATAGGC GCTCAAAACC 300

TATCATTCTA TTCTTTTTCC AGTCTTTGAT ATGGGTTTTA GATTCTTCTA CTTCTTGGAC 360

TTCTAGCTCA TCCGCAAACA AACTCAATTG TTGAGATTGC TTTTGTTTAG CTGAATAAGG 420

ACTACTTTTT TTCAATCCAT CCATCTGAAA GACATTGTAA GAGATAATAG TCGCAATTTC 480

TTTCTTTTGC TCTAATGTTG GTTGATTTCC AGTCTTAGCT AGATAATAGT CCTCAAAAGT 540

TGCCAAAAGA TTCTCACGCG CCAAAAGGAG AGAATCTCCT TGATACTCAT AACCATACGA 600

AGCATGATAA GCATCTTTTA CAAGTTTATA AAATGTGACT TCATCTGAAA CCTCACGACT 660

AATCCGTTGC AGTTTTCTAT CAACAAAACC AACTCGCTCA GATAATGGAA TTTCCTCACC 720

AGTTACGGTA TCATATCTCG TTACCATATA AGGTGCTTCA CCACAAGTTA CCTCTAACCA 780

TCGTAAGTCC ACATACTCCT CAAGACTTAA CGAGCCTAAT TTCGATTCTA CATATCCATT 840

TTGCTTTGCG ACCAACCACG TTGGTGTAAA CACTTCTGCC CTTATTTTTG TCCGATCTTT 900

TTGTTCATAT TTGGATTTTT CAGATCTGGG CTGAATCAAG TTGGCAAAGT TTCCAGTAAC 960

CTTACTTGGA TTGATGCGAT CACTTGGAGC AAATCCCTTT CCTAACAATT CATAAGAATG 1020

CGTAnGCCAA ACAATTGATT TCTTTGTCGT TCGATCTTTT AAAAGAATTT TTAATAAGTC 1080

AGCCGATTCT TTAGCCAAAC TTTCTTCACT AATATCTATT GTCATCAGCA ACCTCTCTTA 1140

TATTGTAAGC CCTATTATAT CATATTTTAA AGAATGAAAA TTTACTTGAA AAAAGTAATT 1200

CAATAAATAT CTCTCCGATG ACCAACTTCT AGAGTAGCAA CGACTAATTC ATCATCTACA 1260

ATTTGTACGA TAACTCGATA ATTACCAATT CTATAGCGCC ATTGACCAAC GCGATTACCA 1320

ACCAAAGCCT TTCCGTGTCG TCTTGGGTCT TCCAAAACAT TGGTTTGTAA ATAGTTTGTA 1380

ATTAGCTTCT GCGTATAACG GTCCAATTTT TTCAATTGCT TGATAAAACG TCTTGTTGGA 1440

ACTAATTTAT ACAAATTATT CATCCTTCAA GCCTAAATCA TGCATCATTT CTTCCCAAGT 1500

AATGGGTTCA ACTCCTTTTT CCAAGTCTTC TAAATACTCT TGATAGGCTA AATCTGCCAC 1560

ACGAGCATCG TATTCATCTT CTAGGGCTTC AAGAGTTTTG GTGCGAATAA GTTCCGAAAG 1620

GGAAACTCCT TCAAACTTAG CCATTGCTTT CATAAATGTT TTATCAGCTT CAGAAACTTT 1680

TAATGTAATA GTAGTCATCT TTTGTGCTCC CTTTTTTAAT GGTAACACCA TTGTATTACT 1740

TTTTAGGTGT TCAGTCAATA TAAAAAGAAC ACCTTCTCAG CGTTCTTTCT ATATCTCTGT 1800

CAATGGTGTT GCGGTATCTG GTGAGGTATC ATAAACCTTA AAGTCTACTC CGACTCCCAG 1860

ATCAGCTTGA GCCAGCTGAT TGACCATGGT CATATGAGCC AGTTCCTTGA TATTGTTTTC 1920

CTTAGATAAA TGCCCAAGGT AAATCTTCTT AGTACGATTT CCTAGCGTCC GAATCATAGC 1980

TTCAGCACCG TCCTCGTTAG AAAGGTGACC AAGGTCAGAT AGGATTCGTT GTTTGAGTCG 2040

CCAAGCGTAA GAACCTGATC GCAAAATCTC TACATCATGG TTGGCCTCGA TAAGATAACC 2100

ATCCGCATTT TCGACAATGC CCGCCATACG GTCACTGACA TAACCTGTAT CTGTCAAGAG 2160

GACAAAACTC TTATCATCCT TCATAAAGCG ATAGAACTGC GGTGCGACTG CATCATGGCT 2220

TACACCAAAA CTCTCGATGT CGATATCTCC AAAGGTTTTG GTTTTACCCA TTTCAAAAAT 2280

ATGCTTTTGC GAAGAATCCA CCTTGCCAAG ATATTTACTA TTTTCCATAG CTTGCCAGGT 2340

CTTTTCATTG GCATAAAGAT CCATACCATA CTTGCGAGCC AAAACGCCTA CTCCATGGAT 2400

ATGATCTGAA TGCTCATGGG TAATCAAGAT GGCATCCAGG TCTTCTGGCT TACGGTTAAT 2460

TTCAGCTAGC AGACTGGTAA TTTTCTTGCC AGACAAGCCT GCATCTACTA AAAGCTTCTT 2520

TTTTGAGGTT TCCAGATAAA AAGAATTTCC ACTGGAACCC GACGCTAAAA TACTGTATTT 2580

AAAGCCTATT TCACTCATTC TAGTCTTCTA CTTCATCCTC CCATACTTCT TCTTTCACTG 2640

CATCCTTATC ATAAGGGAGT ACAATGGTAA AGGTTGAACC CTTGCCGTAT TCACTCTTGG 2700

CCCAAATAAA GCCCTTATGT TGTTTGATAA TTTCTTTAGC GATAGACAGT CCTAGACCTG 2760

TACCACCTTG TGCACGACTT CTAGCACGAT CCACACGATA GAAACGGTCA AAGATACGTG 2820

GTAAATCCTG CTTAGGAATC CCCAAACCGT GGTCAGAAAT GGATAAAATC ATCTGGTCTT 2880

CAGTTGTCTT CATTCTGACA GTGATTTTAC CCCCATCTGG CGAATACTTA ATAGCATTAT 2940

TTAAAATATT GTCGACAACC TGCGTCATCT TATCTGTATC AATTTCCATC CAGATAGAAT 3000

TGATGGGATA ATCTCTCACC AACTCATATT TTTTCTCCTT TTCCTGTCCT TTCATCTTGT 3060

CAAAACGATT GAGGATAAAG GTAATAAAAG CAGTGAAGTT AATCAGTTCC ACATCTAGGT 3120

GACTGGTAGC ATTATCAATA CGTGAAAGAT GGAGGAGATC CGTCACCATG CGCATCATAC 3180

GGTTGGTCTC ATCAAGAGAA ACCTTGATAA AGTCTGGTGC TACAGTTTCA CACAAAGCCC 3240

CCTCATCCAA GGCTTCAAGA TAGGATTTTA CGCTAGTCAG AGGAGTCCGT AACTCATGGC 3300

TAACATTGGA AACAAAGAGT CTTCGTTCGC GTTCTTCCTT CTCCTGCTCC GTCGTATCAT 3360

GCAAAACAGC CACCAAACCT GAAATAAAGC CAGACTCTCG ACGTATCAAG GCAAAGCGAA 3420

CTCGAAGGTT CAAATATTCG CCATTGATAT CTTGGGAATC TAGCAACAAT TCTGGACTTT 3480

GGGTAATCAA ATCACGCAAT TCATAGTTTT CTTCTATCTT GAGCAATTCC AAAATGCTTC 3540

TATTCAGAAC ATCTTCCTTA ACCAACCCCA GTTGCTTCTT GGCTGTATCG TTAATCATGA 3600

TAATCTGACC CCGACGGTTA GTCGCAAGAA CCCCATCTGT CATATAAAAC AGAATACTAT 3660

TTAGCCTCTT ACTCTCTTGT TCTAGATTTT CCTGAGTGAG ACGAATAACC TCCGACAAGT 3720

CATTCAAATT ATTGGTAATA TTGGTGATTT CAGACCCACC TTGCATATCA AGAACCTTGG 3780

AATAATCTCC TGCAATCAAA TCTTTAACCT TTTGATTGAC TTGCTTCAAC TGAATATTAT 3840

CACGTCTATT TTCCAGTAAT AAGAGGGTCA CAACAAGGAT GAAACCTAAC AAAATCAGGA 3900

TAAAGATAAA ATCTCTGGTA AAAATGGTTT GTTTCAGTAA ATCAAGCATT ATTTCTCATG 3960

TAATACCCTA CACCACGGCG CGTCAAGATA TACTCTGGTC GGCTGGGCGT ATCTTCAATC 4020

TTCTCACGCA GACGTCGTAC AGTCACATCA ACTGTACGGA CATCACCAAA ATAGTCATAA 4080

CCCCAGACAG TCTCAAGCAA GTGTTCGCGC GTGATGACTT GACCTGTATG CGATGCTAAA 4140

TGATACAAAA GCTCAAATTC ACGATGGGTT AAGTCTAGTT CTTCGCCATA TTTTTTAGCC 4200

ACGTAGGCGT CTGGAACAAT TTCTAAATCC CCAATTTGGA TAGGTTGAGG TTTACTATCT 4260

GCTTCCTGAC CATCTACTGG CATAGGTTGA GAACGACGCA GAAGAGCTTT AACACGCGCC 4320

TGCAACTCAC GATTGGAGAA GGGTTTTGTT ACATAGTCAT CTGCCCCAAG TTCCAAACCG 4380

ATAACCTTAT CAAATTCACT ATCTTTGGCT GAAAGCATAA GAATGGGCAC ACTGCTTGTC 4440

TTACGAATGG TCTTAGCAAC TTCTAAACCA TCAATTTCTG GAAGCATCAA ATCCAGAATA 4500

ATAATATCTG GTTGCTCTGC TTCAAATTGC TCTAGCGCTT CACGACCATT AAAAGCAGTT 4560

ACAACTTCGT AACCTTCCTT GGTCATATTA AACTTGATAA TATCCGAGAT TGGTTTCTCA 4620

TCATCTACAA TTAGTATTTT TTTCATATGT TCACCTTTTT CTCTACTATT ATACCAAAAA 4680

AATAGTCAGA AGACACAATA GCTAGTCTTG GCTACTGTCT AAGTTGGCTT GTGCATAAAC 4740

CTGCCAGATT TTTTGTTGGG GTTTGGCAAG TGGGTAATTC TTGAATTCTT CTGGTGAAAG 4800

CCAGCGAACT TCCCTATCTG AAAAATCATG GAAGTCACTC ACCTGACCTG CTACAATCTG 4860

TACATGCCAT TTTCGATGAC TAAAAACATG CTGGACTGTA TCAAAACAAA CATCAAGCCA 4920

ATCAACATCT AGGTCATAGT CCTGCTGGAA ACTCTCTTCT GGACTGGGAC CAAAGTTCAC 4980

ACTTTCTTCC GCAACCTGAT GAAAGAGGTC AAACTGCTCT TCTTGCGAAA AGTTATCAAC 5040

TTCTATAAAG GGGAAATGCC AAAAACCTGC CAAGAGCTTT TCGCTTTCAT TTTTTTCAAG 5100

TAAAAATTGT CCTTGAGAAT TTTTCACAAC TAAGGCTTTA AGATAAATAG GAACCGGCTT 5160

TTTCTTAGGA GATTTAATTG GATAACGGTC CATGGTTCCA TTCTGATATG CCGCACTAAA 5220

GTCCTTGACT GGGCTTTCTT CAGGTCTGGG ATTTACAGGA GACTCAATAT CAGACCCTAA 5280

GTCCATCAAG GCTTGATTAA AATCACCCGG ACGATCCGGA TTAATCAAGA TCTCCATCAT 5340

TGCCTGAAAA ATTTTTCGAT TACTTGGAAT CCCAATATCG TGGTTGACTT CAAACAGACG 5400

CGCCAAGACC CGCATGACAT TACCATCTAC AGCTGGCTCA GGCAAGTTAA AAGCAATACT 5460

GGAAATGGCT CCTGCTGTGT AAGGTCCAAT CCCTTTCAAG CTGGAAATTC CTTCATAGGT 5520

ATTTGGAAAT TGGCCACCAA AGTCAGTCAT AATCTGCTGG GCTGCAGCCT GCATATTGCG 5580

AACTCGAGAA TAATAGCCCA AGCCCTCCCA AGCTTTCAGT AAACTCTCCT CAGGCGCAGT 5640

TGCCAGACTT TCGACAGTTG GAAACCAGTC CAAAAATCTT TCGTAGTAAG GGATAACTGT 5700

ATCCACCCTG GTCTGCTGAA GCATGATTTC AGATACCCAG ATGTGATAAG GATTTTTACT 5760

TCTCCTCCAA GGCAAATCTC TTTTGTTTTC ATCATACCAA GCGAGAAGTT TCTCACGGAA 5820

AGAAATGACT TTCTCCTCCG GCCACATGAC GATACCGTAT TCTTTCAAAT CTAACATATC 5880

TCTAGTATAA CACAGAAGGT TTCACCTGTC TTTGTATCTG ATTTATAATA TTTTCAATAG 5940

ATAGTATATA ACTTTTCTAT CTACTTATAC TCAATGAAAA TCAAAGAGCA AACTAGGAAG 6000

CTAGCCGCAG GTTGCTCAAA ACACTGTTTT GAGGTTGTGG ATAGAACTGA CAGAGTCAGT 6060

ATCATATAcT ACGGCAAGGT GAAGCTGACG TAGTTTGAAG AGATTTTCGA AGAGTATAAA 6120

TCTTATTGAT GAACTGCTTG CAGTCTGAGA AAAAATGAGC TTGGATATTA TTTCCAAACT 6180

CACTTAAAGT CAATTTCAAT CCACTAGAAC AAGCCTAGTA CAGTTCCATC GCTTTCAACA 6240

TCCATGTTGA GAGCTGCTGG ACGTTTTGGA AGACCTGGCA TGGTCATAAC ATCACCAGTT 6300

AAGGCAACGA TGAAGCCTGC ACCTAATTTT GGTACCAATT CACGAATGGT AATTTCAAAG 6360

TTTTCTGGTG CTCCAAGCGC ATTTGGATTG TCTGAGAAAC TGTATTGAGT TTTAGCCATA 6420

CAGATTGGCA ATTTGTCCCA ACCGTTTTGA ACGATTTGAG CAATTTGTGT TTGAGCTTTC 6480

TTCTCAAAGT TCACTTTGCT ACCACGATAG ATTTCAGTGA CAATTTTTTC AATCTTTTCT 6540

TGGACAGAAA GGTCATTATC ATACAAACGT TTATAGTTAG CTGGATTTTC AGCAATTGTC 6600

TTAACAACTG TTTCGGCAAG TGCTACTCCA CCTTCTGCTC CATCAGCCCA GACACTAGCC 6660

AATTCAACTG GTACATCGAT TGAGGCACAG AGTTCTTTTA AGGCTGCAAT TTCAGCTTCT 6720

GTATCAGATA CAAATTCGTT AATAGCTACA ACTGCTGGAA TACCGAACTT ACGGATATTT 6780

TCAACGTGGC GTTTCAAGTT AGCAAAACCT GCACGAACTG CCTCTACATT TTCTTCAGTC 6840

AGAGCGTCTT TAGCCACACC ACCATTCATC TTAAGGGCAC GAAGGGTTGC GACAATAACA 6900

ACTGCATCTG GAGATGTTGG CAAGTTTGGT GTCTTGATAT CAAGGAATTT CTCAGCACCA 6960

AGGTCCGCAC CAAAACCAGC TTCAGTAACA GTGTAATCAG CCAAGTGAAG GGCTGTTGTC 7020

GTCGCCAAAA CAGAGTTACA GCCATGAGCG ATATTGGCAA ATGGACCACC GTGTACAAAG 7080

GCAGGTGTAC CGTAAATTGT CTGAACCAAG TTTGGCTTAA TAGCATCCTT CAAAATCAAA 7140

GCCAAGGCAC CCTCAACCTG CAAATCACCT ACAGAAACAG GCGTACGGTC ATAGCGATAA 7200

CCAATAACGA TATTCGCCAA ACGACGTTTC AAGTCCTCGA TGTCCGTTGC CAAGCAAAGA 7260

ATTGCCATGA TTTCTGAAGC AACTGTAATA TCAAAACCAT CCTCACGTGG AATACCGTTT 7320

AGAGGACCAC CAAGACCAAC AGTCACATGG CGGAGCGTAC GGTCGTTCAA GTCCACAACG 7380

CGTTTCCAGA GGATACGACG TTGATCAATT CCCAGCTCAT TCCCTTGGTG CAAGTGGTTG 7440

TCAATCAAGG CAGAAAGGGC ATTGTTGGCA GTTGTAATAG CATGCATATC TCCAGTAAAG 7500

TGGAGGTTGA TGTCTTCCAT TGGCAGAACT TGTGCATACC CACCACCAGC AGCACCACCC 7560

TTGATCCCCA TGACTGGACC AAGAGACGGT TCGCGGATAG CAATCATGGT TTTCTTGCCA 7620

ATCTTGTTCA AGGCATCCGC AAGACCAATG GTAAGCGTCG ACTTTCCTTC ACCTGCAGGT 7680

GTTGGGTTGA TGGCAGTAAC CAAGATCAAT TTACCGACTG GATTGCTCTC AACTGCACGA 7740

ATTTTATCAA AGCTGAGTTT AGCCTTGTAC TTTCCGTACA ACTCCAAATC GTCATAAGAA 7800

ATACCAAGTT TCTCTACAAC ATCAACAATT GGCTTCAACT CAATACTCTG TGCGATTTCA 7860

ATATCTGTTT TCATTCAAAA TTCCTCTAAC CTCTTATATG ATAATTCATT ATATCACAAA 7920

ACAAGATTTT TAACATCCTA AAACTCTCTA AACGTTCGTA AATATCTCTG TTTTTAAGAC 7980

TTTTAGAGTC CTTTCTTAAA TTTTATATGG CTTTATAGTT TGAAACTATA ATAAATCTTC 8040

GTTTTTACCA AAAATTTATC ACTTTCATTT TACTTACCGC TTATTTTTGT GTACAATAGT 8100

GCTATGAAAA TTTTAGTTAC ATCGGGCGGT ACCAGTGAAG CTATCGATAG CGTCCGCTCT 8160

ATCACTAACC ATTCTACAGG TCACTTGGGG AAAATTATCA CAGAGACTTT GCTTTCTGCA 8220

GGGTATGAAG TTTGTTTAAT TACGACAAAA CGAGCTCTGA AGCCAGAGCC TCATCCTAAC 8280

CTAAGTATTC GAGAAATTAC CAATACCAAG GACCTTCTAA TAGAAATGCA AGAACGTGTT 8340

CAGGATTATC AGGTCTTGAT CCACTCAATG GCTGTTTCTG ACTACACTCC TGTTTATATG 8400

ACAGGGCTTG AGGAAGTTCA GGCTAGCTCC AATCTAAAAG AATTTTTAAG CAAGCAAAAT 8460

CATCAGGCCA AGATTTCTTC AACTGATGAG GTTCAGGTTT TGTTCCTTAA AAAGACACCC 8520

AAAATCATAT CCCTAGTCAA GGAATGGAAT CCTACTATTC ATCTGATTGG TTTCAAACTG 8580

CTGGTTGATG TTACCGAAGA TCATCTGGTT GACATTGCAC GAAAAAGTCT TATCAAGAAT 8640

CAAGCAGATT TAATCATCGC GAATGACCTG ACTCAAATTT CAGCACATCA GCACCGAGCT 8700

ATATTTGTTG AGAAAAATCA GCTTCAAACA GTCCAGACTA AAGAAGAAAT TGCAGAACTC 8760

CTCCTTGAAA AAATTCAAGC CTATCATTCT TAGAAAGGAA AACTATGGCA AACATTCTCT 8820

TGGCTGTAAC GGGTTCAATC GCCTCTTATA AGTCGGCAGA TTTAGTCAGT TCTCTAAAAA 8880

AACAAGGCCA TCAAGTCACT GTCTTAATGA CTCAGGCTGC TACAGAGTTT ATCCAACCTT 8940

TGACACTACA GGTACTCTCA CAGAATCCTG TCCACTTGGA TGTCATGAAG GAACCCTATC 9000

CTGATCAGGT CAATCATATC GAACTTGGAA AAAAAGCAGA TTTATTTATC GTGGTACCTG 9060

CAACTGCTAA CACTATTGCA AAACTAGCTC ACGGATTTGC GGACAACATG GTAACCAGTA 9120

CAGCTCTAGC CCTACCAAGT CATATTCCCA AACTAATAGC TCCTGCTATG AATACAAAAA 9180

TGTATGACCA TCCAGTAACT CAGAATAATC TGAAAACATT AGAAACTACG GCTATCAGCT 9240

GATTGCTCCT AAGGAATCCC TACTAGCTTG TGGAGACCAC GGACGAGGAG CTTTAGCTGA 9300

CCTCACAATT ATTTTAGAAA GAATAAAGGA AACTATCGAT GAAAAAACGC TCTAATATTG 9360

CACCCATTGC TATCTTTTTT GCTACCATGC TCGTGATACA CTTTCTGAGC TCACTTATCT 9420

TTAACCTTTT TCCATTTCCA ATCAAACCGA CCATTGTTCA TATTCCTGTC ATTATTGCCA 9480

GCATTATTTA TGGTCCACGA GTTGGGGTTA CACTTGGATT TTTGATGGGA TTACTTAGCT 9540

TGACGGTTAA CACGATTACG ATTCTACCGA CAAGCTACCT CTTCTCTCCC TTCGTACCAA 9600

ACGGAAACAT CTACTCAGCT ATCATTGCCA TCGTCCCACG TATTTTGATT GGTTTAACTC 9660

CTTACTTAGT CTATAAACTG ATGAAAAACA AGACTGGTCT GATTTTAGCT GGAGCCCTTG 9720

GTTCCTTGAC AAATACTATC TTTGTCCTTG GAGGAATCTT CTTCCTATTT GGAAATGTTT 9780

ATAATGGAAA TATCCAACTT CTTCTGGCAA CCGTTATCTC AACAAATTCA ATTGCTGAAT 9840

TGGTCATTTC TGCAATTCTA ACCCTAGCCA TTGTTCCACG ACTACAAACC TTGAAAAAAT 9900

AAAAACAGG 9909

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1126 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TAATTTTCAT ATAATAGTAA AATAGAATGT GTGATTCAAT AATCACCTCA AATAGAAAGG 60

AAATTCTATG TCAAATCTAT CTGTTAATGC AATTCGTTTT CTAGGTATTG ACGCCATTAA 120

TAAAGCCAAC TCAGGTCATC CAGGTGTGGT TATGGGAGCG GCTCCGATGG CTTACAGCCT 180

CTTTACAAAA CAACTTCATA TCAATCCAGC TCAACCAAAC TGGATTAACC GCGACCGCTT 240

TATTCTTTCA GCAGGTCATG GTTCAATGCT CCTTTATGCT CTTCTTCACC TTTCTGGTTT 300

TGAAGATGTC AGCATGGATG AGATTAAGAG TTTCCGTCAA TGGGGTTCAA AAACACCAGG 360

TCACCCAGAA TTTGGTCATA CGGCAGGGAT TGATGCTACG ACAGGTCCTC TAGGGCAAGG 420

GATTTCAACT GCTACTGGTT TTGCCCAAGC AGAACGTTTC TTGGCAGCCA AATATAACCG 480

TGAAGGTTAC AATATCTTTG ACCACTATAC TTACGTTATC TGTGGAGACG GAGACTTGAT 540

GGAAGGTGTC TCAAGCGAGG CAGCTTCATA CGCAGGCTTG CAAAAACTTG ATAAGTTGGT 600

TGTTCTTTAT GATTCAAATG ATATCAACTT GGATGGTGAG ACAAAGGATT CCTTTACAGA 660

AAGTGTTCGT GACCGTTACA ATGCCTACGG TTGGCATACT GCCTTGGTTG AAAATGGAAC 720

AGACTTGGAA GCCATCCATG CTGCTATCGA AACAGCAAAA GCTTCAGGCA AGCCATCTTT 780

GATTGAAGTG AAGACGGTTA TTGGATACGG TTCTCCAAAC AAACAAGGAA CTAATGCTGT 840

ACACGGCGCC CCTCTTGGAG CAGATGAAAC TGCATCAACT CGTCAAGCCC TCGGTTGGGA 900

CTACGAACCA TTTGAAATTC CAGAACAAGT ATATGCTGAT TTCAAAGAAC ATGTTGCAGA 960

CCGTGGCGCA TCAGCTTATC AAGCTTGGAC TAAATTAGTT GCAGATTATA AAGAAGCTCA 1020

TCCAGAACTG GCTGCAGAAG TAGAAGCCAT CATCGACGGA CGTGATCCAG TCGAAGTGAC 1080

TCCAGCAGAC TTCCCAGCTT TAGAAAATGG TTTTtCTCAA GCAACT 1126

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2520 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CCGGCAACAA AAAAGAAAAA ATCAACAGTT AAAAAAAATC TAGTCATCGT GGAGTCGCCT 60

GCTAAGCCAA GACGATTGAA AAATATCTAG GCAGAAACTA CAAGGTTTTA GCCAGTGTCG 120

GGCATATCCG TGATTTGAAG AAATCCAGTA TGTCCGTCGA TATTGAAAAT AATTATGAAC 180

CGCAATATAT TAATATCCGA GGAAAAGGCC CTCTTATCAA TGACTTGAAA AAAGAAGCTA 240

AAAAAGCTAA TAAAGTTTTT CTCGCGAGTG ACCCGGACCG TGAAGGAGAA GCGATTTCTT 300

GGCATTTGGC CCATATTCTC AACTTGGATG AAAATGATGC CAACCGTGTG GTCTTCAATG 360

AAATCACCAA GGATGCAGTC AAAAATGCTT TTAAAGAACC TCGTAAGATC GATATGGACT 420

TGGTCGATGC CCAACAAGCT CGTCGGATCT TGGATCGCTT GGTAGGGTAT TCGATTTCGC 480

CTATTTTGTG GAAGAAGGTC AAGAAGGGCT TGTCAGCAGG TCGCGTTCAG TCCATTGCCC 540

TTAAACTCAT CATTGACCGT GAAAATGAAA TCAATGCCTT CCAGCCAGAA GAATACTGGA 600

CAGTTGATGC TGTCTTTAAA AAGGGAACCA AACAATTTCA TGCTTCCTTC TATGGAGTAG 660

ATGGTAAAAA GATGAAACTG ACCAGCAATA ACGAAGTCAA GGAAGTCTTG TCTCGTCTGA 720

CGAGTAAAGA CTTTTCAGTA GATCAGGTGG ATAAGAAAGA GCGCAAGCGC AATGCTCCTT 780

TACCCTATAC CACTTCATCT ATGCAGATGG ATGCTGCCAA TAAAATCAAT TTCCGTACTC 840

GAAAAACCAT GATGGTTGCC CAACAGCTCT ATGAAGGAAT TAATATCGGT TCTGGTGTTC 900

AAGGTTTGAT TACCTATATG CGTACCGATT CGACTCGTAT CAGTCCTGTA GCGCAAAATG 960

AGGCGGCAAG CTTCATTACG GATCGTTTTG GTAGCAAGTA TTCTAAGCAC GGTAGCAAGG 1020

TCAAAAACGC ATCAGGTGCT CAGGATGCCC ATGAGGCTAT TCGTCCGTCA AGTGTCTTTA 1080

ATACACCAGA AAGCATCGCT AAGTATCTGG ACAAGGATCA GCTTAAGCTA TATACCCTTA 1140

TCTGGAATCG TTTTGTGGCT AGCCAGATGA CAGCGGCCGT TTTTGATACC ATGGCTGTTA 1200

AATTGTCTCA AAAAGGGGTT CAATTTGCTG CCAATGGTAG TCAGGTTAAG TTTGATGGTT 1260

ATCTTGCCAT TTATAATGAT TCTGACAAGA ATAAGATGTT ACCGGACATG GTTGTTGGAG 1320

ATGTGGTCAA ACAGGTCAAT AGCAAACCAG AGCAACATTT CACCCAACCG CCTGCCCGTT 1380

ATTCTGAAGC AACACTGATT AAAACCTTAG AGGAAAATGG GGTTGGACGT CCATCAACCT 1440

ACGCGCCAAC CATTGAAACC ATTCAGAAAC GTTATTATGT TCGCCTGGCA GCCAAACGTT 1500

TTGAACCGAC AGAGTTGGGA GAAATTGTCA ATAAGCTCAT CGTTGAATAT TTCCCAGATA 1560

TCGTAAACGT GACCTTCACA GCTGAAATGG AAGGTAAACT GGATGATGTC GAAGTTGGAA 1620

AAGAGCAGTG GCGACGGGTC ATTGATGCCT TTTACAAACC ATTCTCTAAA GAAGTTGCCA 1680

AGGCTGAAGA AGAAATGGAA AAAATCCAGA TTAAGGATGA ACCAGCTGGA TTTGACTGTG 1740

AAGTGTGTGG CAGTCCAATG GTCATTAAAC TTGGTCGTTT TGGTAAATTC TACGCTTGTA 1800

GCAATTTCCC AGATTGCCGT CATACCCAAG CAATCGTGAA AGAGATTGGT GTTGAGTGTC 1860

CAAGCTGTCA TCAGGGACAA ATTATTGAGC GAAAAACCAA GCGTAATCGC CTATTCTATG 1920

GTTGCAATCG CTATCCAGAA TGTGAATTTA CCTCTTGGGA CAAGCCTGTT GGTCGTGACT 1980

GTCCAAAATG TGGCAACTTC CTCATGGAGA AAAAAGTCCG TGGTGGTGGC AAGCAGGTTG 2040

TTTGTAGCAA AGGCGACTAC GAGGAAGAAA AGATGGCTCT TTGTCAACTG TAGTGGGTTG 2100

AAGTCAGCTA AGCTCGAGAA AGGACAAATT TTGTCCTTTC TTTTTTGATA TTCAGAGCGA 2160

TAAAAATCCG TTTTTTGAAG TTTTCAAAGT TCCGAAAACC AAAGGCATTG CGCTTGATAA 2220

GTTTGATGAG ATTATTGGTC GCTTCCAATT TGGCGTTAGA ATAGTGTAGT TGAAGGGCGT 2280

TGACGATTTT CTCTTTGTCC TTTAGAAAGG TTTTAAAGAC AGTCTGAAAA AGAGGATGAA 2340

CCTGCTTTAG ATTGTCCTCA ATGAGTCCGA AAAATTTCTC CGGTTCCTTA TTCTGAAAGT 2400

GAAACAGCAA GAGTTGATAG AGCTGATAGT GATGTTTCAA GTCTTGTGAA TAGCTCAAAA 2460

GCTTGTTTAA AATCTCTTTA TTGGTTAAAT GCATACGAAA AGTAGGGCGA TAAAAATGTT 2520

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10993 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TTTTCTCGAT AATAACTTCC ACCTTATTAT TTGGGATACC CTCCTCTTCT TCACCACCAC 60

GTTCATAGTA GTCATCGCGA TAGAGAAAAG CTACGATATC AGCGTCCTGC TCAATAGACC 120

CAGATTCACG AATATCAGAC AAGACCGGTC TCTTGTCCTG ACGTTGTTCT ACACCACGAG 180

AAAGCTGACT CAGAGCGATT ACTGGAACCT TCAATTCCTT GGCTAGTATT TTCAACTGAC 240

GAGAAATTTC AGAAACTTCT TGTTGACGAT TTTCTCGACC AGTTCCCGTG ATAAGTTGCA 300

AATAGTCTAT CAAAATCAAA CCAAGATTTC CAGTTTCTTG AGCCAATTTA CGAGAACGAG 360

AACGAATCTC TGTAATCCGA ATACCTGGCG TATCATCGAT ATAGATACTG GCGTTAGCTA 420

GATTACCCTG AGCAATAGTA TATTTTTGCC ACTCCTCATC TGTCAATTGC CCTGTACGGA 480

TAGAATGTGA CTCCACTAAG CCTTCTGCAG CTAACATACG ATCTACCAAG CTTTCCGCAC 540

CCATTTCGAG TGAAAAAATA GCAACCGTTT TGTCCAACTT AGTCCCAATG TTCTGAGCGA 600

TATTCAAGGC AAATGCTGTC TTACCAACTG CTGGACGAGC TGCTAAGATA ATCAACTCCT 660

CCTCATGAAG TCCTGTTGTC ATATGATCCA AATCACGATA ACCTGTCGCA ATACCTGTAA 720

TATCGGTCGT TTGTTGCGAG CGAGCTTCCA GATTTCCAAA GTTGAGATTC AACACATCTC 780

GAATGTTCTT AAACCCGCTT CGATTTGCAT TTTCACTGAC ATCAATCAAC CCTTTTTCTG 840

CCTGAGCAAT AATTTCATCA GCTGGTTGTG ACGCTTCGTA AGCTTGGTTG ACAGACTCTG 900

TCAACTTGGC AATTAAACGA CGTAGCATTG CTTTTTCTGC AACAATCTTA GCATAATACT 960

CCGCATTAGC AGAAGTTGGC ACAGAATTAA CAATCTCAAC CAAGTAAGAC AAGCCACCAA 1020

TATTCTGTAA ATCACCTTGA TTATCAAGGA TAGTACGAAC CGTTGTTGCA TCTATGGCAT 1080

CACCACGATC GGATAAATCG ACCATGGCTT GGAAAATCAA ACGATGGGCA TACTTAAAAA 1140

AGTCCCGAGA CTCAATGTAT TCTCGCACAA AAACAAGTTT ACTCTCATCA ATAAAGATAG 1200

CCCCTAAAAC GGATTGCTCA GCTAAGATAT CTTGAGGTTG TACTCGTAAC TCTTCTACTT 1260

CTGCCATCAG ACTTCCCTTC CTTTTACAAT CTTGTCAAGA AGGTGTAAAC TTATCCTTCT 1320

TTCACACGAA GATTGATTAC ACTTGTGATA TCTTGATAGA TTTTCACTGG CACATCAATC 1380

AAACCAACCG CTCGAATCGG AGCTTGTACT TGAATATGAC GTTTATCAAT CTTAATTCCA 1440

AATTGCTTTT GCAATTCTTC TGCAATCTTC TTATTGGTAA TAGAACCAAA GGTACGACCA 1500

TCTGGACCAA CTTTTTCAAC AAATTCTACA ACAGTTTCTT CTGCTTCAAG TTGTGCTTTA 1560

ATTGCTTTTC CTTCTGCAAT CATCTCAGCG TGAGCTTTTT CTTCCGATTT TTGTTTACCA 1620

CGAAGTTCAC CTACAGCTTG AGCAGTCGCT TCTTTGGCTA GATTCTTTTT GATAAGAAAG 1680

TTTTGCGCAT ACCCTGTTGG TACTTCCTTA ATTTCGCCTT TTTTACCTTT TCCTTTAACA 1740

TCTGCTAAAA AGATTACTTT CATTCTTCTT TCTCCTTTTC CTTCATTTCA TTTAATACAA 1800

TTTCTGTCAG TTTTTCACCT GCTTCTGACA AGGTTACATC TTTAATTTGA GCTGCTGCCA 1860

AATTAAAGTG GCCTCCACCG CCTAACTCTT CCATAATCCG TTGTACATTC AGTTTACTAC 1920

GACTTCGAGC TGAGATAGAG ATAAATCCTT GTGTATTCTT CGCAAGAACA AAACTCGCTT 1980

CAATACCTGA CATGGCTAAC ATGGCATCTG CTGCCTTACT AATAACAACT GTATCATAGC 2040

ATTTCATGTC CTTAGCCTCT GCTATTAGTA CATCTGAACC TAATTTACGC CCCTGTAAAA 2100

TAAGTTCATT GACCTCACGA TATTCTTCAA AATCTGTCGC AGCGATTTCC TGGATAGCAA 2160

TACTATCACT TCCGCGCGTT CTGAGATAGC TAGCAACATC AAATGTCCGA CTAGTTACTC 2220

GCGAGGTGAA ATTTTTAGTA TCCAACATCA TACCAGCCAT CAAGACACTT GCTTGCATAC 2280

GACTCAAACG ATTTTTCTTA GAATTCTGGA ACTGAATCAA TTCCGTTACC AACTCACTGG 2340

CACTACTTGC ACCACTTTCG ATATAAGTAA TAACCGCATT ATCTGGAAAA TCCTGATCCC 2400

TTCTATGGTG GTCAATAACA ATGGTTTGGG TAAATAAATC ATAAAATTCT TTTGATAATG 2460

TTAAGGCTGT CTTTGAATGG TCTACAAGAA TCAACAAAGA ACGATTGGTC ACCATCCCCA 2520

TTGCATCCTT AACAGACAAC AACTTCGTAA CTCCTTCTTT TTCTATGAAT GAAACAGCTC 2580

GTTCAATATC TGGAGACATT TGTTCTTCAT CATAAAGAGC ATAGCTATTT TCAATCACAT 2640

TGCTGGCGAA CAACTGCATA CCTACAGCAG AGCCCAAAGC ATCCATGTCT AAATTTTTGT 2700

GACCGACTAC AAAAACCTGA TCTACACTCC GAATCTTATC TGAAATAGCT GTCATCATAG 2760

CGCGCGTACG AGTCCGTGTA CGCTTGATTG AAGCAGCAGA CCCACCACCA AAATAAACTG 2820

GATTTTTCGT TTCGTCGTTT TCCTTAACAA CCACCTGGTC GCCACCACGT ACTTCAGCCA 2880

AGTTCAAATT GAGCAAAGCA ACTTTCCCTA TCTCATCATG ATTTCCATCG CCATAAGAAA 2940

ATCCCATACT TAAGGTCAAG GGCAACTGTC TCTGTTTCGA CTCTTCTCTG AAAGCATCAA 3000

TAACAGAAAA TTTATCATTC ATCAAGCCCT CAAGCACCGT GTAGTCAGTA AATAGATAAA 3060

ATCGATCCAT ACTTACCCGA CGAGAAAACA TCATGTGTTT TTCTGAAAAC TCTGATATAA 3120

AATTAGCTAC AAAACTATTG ATTTGACTAA TATCTGACTC AGAAGTTTCA TCCTCCAAAT 3180

CATCATAATT ATCCACAGAG ACAATCCCAA TCACTGGTCT ACTTGTTACC AATTCATCTG 3240

TTATGGCTTG TTCCCTGGAT ACATCTACAA AATACAAAAC ACCGGAAGAA GCATCCATAT 3300

GAACAGCATA ACGCTTCTCA CCAAGCTTGG CATAAGTAGA CGGATTTCCT ACTGAAGCCT 3360

TGATAATCGT TTGAACAGCT TCTAAATCAA AATCACCATC TTCCTTGGTC AAAATCAATT 3420

CAGCATAGGG ATTAAACCAC TCAACCTCTC CAGAAGATAA ATTCAATTTC ATAACACCTA 3480

CAGGCATCTG TTCCAATAGA GCTGTCAAAC TTTCTTCCGC TTGGTGGTTT ACATACTGTA 3540

TCTGTTCTAC ATCACTCCTT GTATAATGCA CTCTCAGTTT CTTAAATAAA AAAACATAGC 3600

CTCCTACAAA AAGAAACAAA ATTAAAACCG TCAACAGATT ATTATTAACA AAAATAATGA 3660

AAGTGGATAA GACTCCAAAC GCAATCAATC CTACTAGAAT AGGAAAAATT GGACTTACAT 3720

AAAATTTTTT CATTCAAAAC CTCTTGGCAC CCATTATACC ATAATACCCC TCAAAAAGCG 3780

ACTTTTTAAA AGTGTAATCA GTAATTCTAT CAATTATAAG AAAAAGGTAG TTTACAATTC 3840

AGTAAACCTA CCTTTACACA TATTGAAATT AAGATTCTTT AACCTCTAAC AAACCAATTT 3900

CGCCATCCTC ACGACGATAA ATCACATTGG TTGTCTGATC TTCAACATCC ACATAGATAA 3960

AGAAATCATG CCCCAATAAA TCCATTTGTA GAATTGCTTC TTCCAAATCC ATTGGTTTTA 4020

AATCAATTTG TTTTGAACGA ACAACTTTAG ACTGGACAAT ATTTGAATCT TCCACCAAAG 4080

CATCTGTAAA TAATTGACCA GTTGCTACCT TATTTTTATT TTTACGCTCG ATTTTTGTTT 4140

TATTTTTACG AATCTGACGT TCAATTTTAT CAGTTACAAG GTCAATTGAA CCATACATAT 4200

CTTGAGATAC ATCTTCTGCG CGGAGAGTAA TAGATCCAAG CGGAATCGTT ACTTCCACTT 4260

TAGCCGTTTT TTCACGATAA ACTTTTAAGT TAATTCGGGC ATCCAACTCT TGTTCTGGTT 4320

GGAAGTACTT TTCGATCTTT TCGAGTTTAG AAACTACATA ATCACGAATT GCTTCTGTTA 4380

CTTCTAGGTT TTCACCACGG ATACTATATT TAATCATATG AGTACCTTCT TTCTAAACAT 4440

TTTTGTTTTT ATGATTTTAT TATAACGCTT TCATTCTATT TTTGCAAATT TTTTCCTCAT 4500

CTTACAAGGG AAAATGTTTT TACATCCTTA GCACCAGCTT CTTCCAACAG TTTCTTAACA 4560

CGATTTATAG TTGCTCCTGT AGTATAGATA TCATCTATAA GTAGGATTTT TTTAGGAATA 4620

GTGACTCCAC TTTTAATAAA GAAAGGAAGT TCTGTCCCCA AGCGCTCTGA ACGATTTTTA 4680

GAAGAACTGG CTCTCTCTTC TCTTTTCTCT AATAAATCCA GATACTCAAA GCCTGCTGCC 4740

TCTACCAAGC CCTCAACCTG ATTAAATCCT CTATTAGCAT ATCTATCAGG ACTTAGGGGA 4800

ATTACAACAA ATTGATACTC TTTGTACTTT TTCAACTCCT CACTTAAAAA TGAAGCGAAA 4860

ACTTTTCTTA ACAGGAAGTC TCCATCAAAC TTATACCGAC TGAAAAAATC CTTCATAGCT 4920

TGATTGTAAG TAAAAATCGC TCTATGACTG ACTTCAACTC CCTCTTTACA CCAAAGTTGA 4980

CAATCTTGAC ACTTTGTTGA CAACTCTGTT TTCATACAAT TTGGACAGTT CTCTTCCCCA 5040

ATTCTTTCAA AAGTAGAATC ACAGTCTGAA CAAAGACAAG AGTCATCATT CCTCAGAAGT 5100

AAGAGACTAC TAAAAGTTAA AACAGTCTTC ATAGTCTGCC CACATAACAA GCACTTCATA 5160

GACCAGCCTC CTTATTCATC ATCTGAATTT CCTTAATCGC CTTCTTGATT GAAGCATTTA 5220

ACCCATCATG GAAGAAAAGC AAATCTCCTG TCGGTCTATC CATGCTTCGT CCAACTCGTC 5280

CACCAATCTG AATCAAACTA GACTTGGTAA ACAAACGATG ATTGGCCTCT ACTACGAAAA 5340

CATCCACACA AGGGAAGGTA ACTCCGCGCT CCAAGATTGT CGTACTGATA AGTATTGTCA 5400

GTTCTCCATC TCGAAAAGCT TGTACTTGCT CTAATCGATC CTCTGTTACA GAAGATACAA 5460

AGCCAATTTT CTCATTTGGA AATTGCTCCT GTAAGATTTC TGCTAACTGC TCCCCTTTCT 5520

TAATTTCTGA AGCAAAAATG AGTAACGGAT AAGCTGTCTT TCTCTGCTTC TCAATATAGG 5580

ACTTTAACTT TGGTGACAAA CGATTCTTGT CTAAGTAGCG ATTAAAATCC GATAACCAAA 5640

TTGGTTTTGG AATAATCAAC GGATTTCCAT GAAACCGTCT CGGTAAATTC AGTCTTTTTA 5700

GTTCTCCTAA ACGGACCTTT TTATCTAACT CATTGGTCGA AGTCGCTGTT AAAAAGATTC 5760

TCAATCCATT CTCCTTTACA CTATTCTTGA CAGCGTGGTA AAGCATGGGA TTATCAACAT 5820

AAGGAAAAGC ATCTACTTCA TCCACTATCA GCAAATCAAA AGCTTGATAA AACTTCAATA 5880

ACTGATGGGT TGTTGCAACA ACTAGTGGTG TTCGAAAATA AGGTTCCGAT TCTCCATGTA 5940

GCAAAGCTAT CCCGCAAGAA AAATCCTGTT GCAGGCGCTT GTACAGCTCC AAACAAACAT 6000

CTATGCGAGG ACTAGCCAAA CACACTGCAC CACCCGCATT GATCACTTTA GCCACTACTT 6060

GATAAATCAT TTCTGTCTTT CCAGCTCCTG TTACCGCATG AACTAAGGTT GGCTTTTGCT 6120

TGTCTACTAC TTGAAGCAAT CCCTCTGACA CCTTCTCTTG AAAAGGAGTT AATTGGCCGC 6180

GCCATTTGAG AACATCTTGC TTTGGAAAAT CCTCCTCCGG AAAATAGTAT AAAGTTTGAT 6240

CACTTCTGAC TCGCTTCATC AGCAAGCACT CTCGACAATA GTAAGCACCG ATGGGCAAAT 6300

ACCATTCTTC TAGAATAGTA CTATTACAGC GTTGACAGAA AAGTTTCCCC TTCTCCTTTC 6360

TCATTGCTGG AAGTTTCTCC GCCAACTGAC GTTCTTCTTC TGTTAATTCA TTCTCAGTAA 6420

ATAAACGACC GAGATAATCT AAATTTACTT TCATACTTCT TTATTCGTAA AAACTAGCAC 6480

TTTAGATGAT TTTTTAGTAC AATTAAATCA TGGAATTTAG GACAATTAAA GAGGACGGTC 6540

AAGTCCAAGA AGAAATCAAA AAATCTCGCT TTATCTGCCA TGCCAAGCGT GTTTATAGCG 6600

AAGAAGAGGC TCGTGACTTC ATTACTGCCA TCAAAAAAGA ACACTACAAA GCGACACATA 6660

ACTGCTCTGC CTTCATTATT GGAGAACGTA GTGAAATTAA ACGTACAAGT GATGATGGTG 6720

AGCCTAGTGG TACTGCTGGT GTTCCCATGC TTGGGGTACT AGAAAATCAC AATCTCACCA 6780

ATGTCTGTGT GGTCGTGACA CGCTACTTTG GTGGTATTAA ACTAGGCGCT GGAGGACTAA 6840

TTCGTGCTTA CGCCGGCAGT GTCGCCTTAG CTGTCAAAGA AATTGGTATT ATTGAAATAA 6900

AAGAACAGGC TGGCATTGCT ATTCAAATGT CTTATGCTCA GTACCAAGAG TACAGTAACT 6960

TCCTTAAAGA ACATGGTCTC ATGGAGCTGG ATACAAACTT TACAGATCAA GTCGATACGA 7020

TGATTTATGT TGATAAAGAA GAAAAAGAAA CTATTAAAGC TGCACTTGTG GAGTTTTTTA 7080

ATGGAAAAGT CACTTTAACT GACCAAGGTT TACGAGAGGT TGAAGTTCCT GTAAACTTAG 7140

TGTAAACAAT GAATAATACA GCGTTTCGTT GACATTCTCA CAACTACTTT AGCGAGCAAA 7200

ATAAAAAAGAG GCGTACCAAA ATATACTAGA AAATGAAGCA ATTCAAACGA AACCTGATAT 7260

CGTTTTCCTT CACACCTATT TACTAGAATT AGCTGAACGC AATCACTTGA AAATTAATGA 7320

CTTTGATCTA TGATATATAG AAATGGTATG GATAGCGTTA TACTAAAGAT ATCTTATACA 7380

AAGAGGTATT CATATGTCTA TTTATAACAA CATTACTGAA TTAATCGGTC AAACACCGAT 7440

TGTTAAACTT AACAACATCG TGCCAGAAGG TGCTGCAGAC GTCTATATAA AGCTTGAAGC 7500

ATTTAATCCT GGTTCATCTG TAAAAGACCG TATTGCCCTT AGCATGATTG AAAAAGCTGA 7560

ACAAGATGGT ATTCTGAAAC CTGGTTCTAC TATTGTTGAA GCAACAAGTG GAAACACCGG 7620

TATTGGACTT TCATGGGTAG GTGCTGCTAA AGGGTATAAA GTCGTCATCG TTATGCCTGA 7680

AACTATGAGT GTAGAACGAC GTAAAATTAT CCAAGCTTAT GGTGCTGAAC TCGTCCTAAC 7740

TCCTGGTAGC GAGGGAATGA AAGGTGCTAT TGCTAAGGCT CAAGAAATCG CTGCTGAACG 7800

TGATGGTTTC CTTCCTCTTC AATTTGACAA TCCAGCTAAT CCAGAAGTAC ACGAAAGAAC 7860

AACAGGAGCT GAGATACTAG CTGCTTTCGG TAAAGATGGA TTAGATGCCT TTGTTGCTGG 7920

AGTAGGTACT GGTGGAACGA TTTCTGGTGT TTCTCATGCA CTCAAATCAG AAAATTCTAA 7980

CATTCAAGTT TTTGCAGTAG AAGCAGATGA ATCTGCTATT CTATCTGGTG AAAAACCTGG 8040

TCCTCACAAA ATTCAAGGTA TCTCAGCTGG ATTTATTCCT GATACACTTG ATACTAAAGC 8100

CTATGATGGT ATCGTTCGTG TAACATCAGA TGACGCTCTT GCACTCGGAC GTGAAATTGG 8160

TGGAAAAGAA GGCTTCCTTG TAGGGATTTC CTCAGCTGCA GCTATCTACG GAGCCATCGA 8220

GGTTGCCAAA AAATTAGGTA CAGGTAAAAA AGTCCTTGCC CTAGCACCAG ATAACGGTGA 8280

ACGTTATCTC TCTACAGCAC TTTATGAATT GTAACCGTCC AATAACGAAG TCTATTGAAA 8340

AATCTCCAGA CTAGAGAACT CACGGATAGT TCCTAATCTG GAGATTTCTT ATTTGCACTT 8400

TTCTTGTACA ACTTTAGTCC ATGGTAAATA GGCCTCTAAA ACCTCTTTGT TTACGAGAGT 8460

TTCCACGTTT GGAAGACATT CTAGAAGATA GGATAGATAT TTCTCACTAT TTATAATGGA 8520

TTGAAATAAG ATATGAACAA ATCGATTAGA ACATGATGGT AAAGCGTAAT CCCTTGTTTC 8580

TCAGCTTTCC CAGACAAAAA AGTCCAATAG TAAGTCAGCT GACTATCACT CTCTAGCACC 8640

CTATAAGAAG TTTCATCCGC ATGAAGTAAG GGCTGAGTCA ATAGTCTCTC TCGCAAGAGG 8700

TTATAAAGGG GCTCCAAATA GTATTGACTC GTCTTGATAT GCCAATTAGA GATTTCCTTA 8760

CGTGTGATTG GTAAACCCAT CCTAGCCCAA TCTTCTTCTT GGCGATAATT GGGTACCTTC 8820

AGATTAAACT TCTGATGGAT GGTGTGAGCG ATAATAGAAG CTGAGCCAAA GTTATGCGCT 8880

AAAGGGGCTT TAGGAATAGG AGCTTTCACA AGCTTATCCA GATGATTATC TTTTACTCGT 8940

TATGGACAAT GCTATATGGC ATAAATCAAG TACCTTAAAG ATTCCGACTA ATATTGGCTT 9000

TGCATTTATT CCTCCATACA CACCAGAGAT GAACCCCATT GAACAAGTGT GGAAAGAGAT 9060

TCGTAAACGT GGATTTAAGA ATAAAGCCTT TCGAACTTTG GAAGATGTCA TACAAGGACT 9120

GGAGAAGGAG GTGATAAAGT CCATCGTTAA TCGGAGACGG ACTAGAATGC TTTTTGAAAA 9180

CAGATGAGTA TAAAAAGAAA GTCCTCATTT CAATAGAAAT CACGACTTTC TGATGAATTT 9240

ATAGTAAAAT GAAATAAGAA CAGGATAGTC AAATCGATTT CTAACAATGT TTTAGAAGCA 9300

GAGGTGTACT ATTCTAGTTT AAATCCACTA TATTTGGGGA GTGATAGAAA AGCCCTTCAT 9360

CAGCCAATCT ACTTGTTCAG GTGCGAGAGC TTTGACATCC TTTTCTGTAC TGGACCAAGT 9420

CAGTTTTCCG TTCTCAAAGC GTTTATATAA TATCCAAAAT CCTTGACCAT CCCAGTAAAG 9480

AACTTTAAAG CGGTCTTTAC GTCCACCACA AAAGAGAAAG ACTTGATCGG AGAAAGGATC 9540

CAATTCAAAG TGGGTTTTAA CTACATAGGC TAATGAGTCT ATTCCCTGCC TCATATCTGT 9600

CTTGCCACAA ACAAGGTGAA CTTGACCTAA ATCACTTAGT TGAATTATCA TAGTACAATA 9660

CCTTTCCTCC GATAATTATT TTTTATCTGG TATACTGGAA GTTGGGGAAT TAGGATAGAT 9720

ACCTTGTTAT GACGCGCTTA CTATGAATTT GAAGTATAGT CTCCTAAATG CACTTAGCCC 9780

TTATTATAGG GCTTTTTGTT TTAATTATTC TAATCGAGTG AGACTGGGGA AAAAACAATT 9840

TCAGGAAAAA TCTAAGCCCT ATACAAAAAA GGAAGCAATT TGCTTCCTTT CTATTATTAG 9900

TTATTCAAGG CTGCTGCCAT TGTAGCTGCA ACTTCAGCTT CGAAGTCGTT TGCAGCTTTC 9960

TCGATACCTT CACCAACTTC AAAGCGAGCA AACTCAACTA CCGAAGCGTT AACTGATTCA 10020

AGGTATGCTT CAACTGTCTT GCTGTCATCC ATGATGTAAA CTTGTGCAAG AAGTGTGTAA 10080

GCTTGGTCAA CTTTAGTGTT ATCAAGCATG AAGCGATCCA TTTTACCTGG AATAATTTTG 10140

TCCCAGATTT TTTCTGGTTT GCCTTCTGCA GCCAATTCAG CTTTGATGTC AGCTTCAGCT 10200

TGAGCAATAA CATCATCAGT TAATTGAGCT TTTGATCCAT ACTTCAAGTG TGGAAGAGCT 10260

GGTTTATTAA CCATTGCACG GCTTTCGTTG TCTTGGTCGA TAACGTGATT CAATTGTGCC 10320

AACTCATCTT TAACGAATTG CTCATCCAAT TCTTTGTAAG AAAGAACTGT TGGTTTCATC 10380

GCTGCGATGT GCATTGACAA TTGTTTAGCA AGTGCTTCGT CTCCACCTTC AACAACTGAA 10440

ATAACACCGA TACGTCCACC GTTATGTTGG TATGCTCCAA AGTGTTGTGC GTCTGTTTTT 10500

TCAATCAATG CAAAGCGACG GAATGAGATT TTCTCTCCGA TAGTTGCTGT TGCAGATACG 10560

TATGCAGCTT CAAGAGTTTC ACCTGAAGGC ATTATCAAAG CAAGAGCTTC TTCGTTGTTA 10620

GCAGGTTTTC CTTCAGCAAT GACTTTAGCT GTAGTATTTA CCAATTCAAC GAATTGAGCG 10680

TTTTTTGCAA CGAAGTCAGT TTCAGCGTTT ACTTCAATAA CTGCTGCAAC ATTACCGTTA 10740

ACATAAACAC CAGTCAAACC TTCTGCAGCA ACACGGTCAG CTTTCTTAGC TGCCTTAGCC 10800

ATACCTTTTT CACGAAGCAA TTCAATCGCT TTTTCGATGT CACCGTCTGT TTCTACAAGC 10860

GCTTTTTTAG CGTCCATAAC ACCGGCACCA GATTTTTCAC GCAACTCTTT TACAAGTTTA 10920

GCTGTAATTT CTGCCATTTT AATTCTCCTA TATTTTTTGA AAATAGGAGA GCGCGGCTAA 10980

GCCCCGCCTC CGG 10993

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8411 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

CGACGGGGAG GTTTGGCACC TCGATGTCGG CTCGTCGCAT CCTGGGGCTG TAGTCGGTCC 60

CAAGGGTTGG GCTGTTCGCC CATTAAAGCG GCACGCGAGC TGGGTTCAGA ACGTCGTGAG 120

ACAGTTCGGT CCCTATCCGT CGCGGGCGTA GGAAATTTGA GAGGATCTGC TCCTAGTACG 180

AGAGGACCAG AGTGGACTTA CCGCTGGTGT ACCAGTTGTC TTGCCAAAGG CATCGCTGGG 240

240

TAGCTATGTA GGGAAGGGAT AAACGCTGAA AGCATCTAAG TGTGAAACCC ACCTCAAGAT 300

GAGATTTCCC ATGATTATAT ATCAGTAAGA GCCCTGAGAG ATGATCAGGT AGATAGGTTA 360

GAAGTGGAAG TGTGGCGACA CATGTAGCGG ACTAATACTA ATAGCTCGAG GACTTATCCA 420

AAGTAACTGA GAATATGAAA GCGAACGGTT TTCTTAAATT GAATAGATAT TCAATTTTGA 480

GTAGGTATTA CTCAGAGTTA AGTGACGATA GCCTAGGAGA TACACCTGTA CCCATGCCGA 540

ACACAGAAGT TAAGCCCTAG AACGCCGGAA GTAGTTGGGG GTTGCCCCCT GTGAGATAGG 600

GAAGTCGCTT AGCTTTAATC CGCCATAGCT CAGTTGGTAG TAGCGCATGA CTGTTAATCA 660

TGATGTCGTA GGTTCGAGTC CTACTGGCGG AGTAATtGAT AAAAGGGaAC ACAGCTGTGT 720

TCCTCTTTTT GTATCAATTT GTATCACCAA GCATTTTCAT AAGGAAGTCT GTTATTTCTT 780

GAGAACTTTC TTTTTTTCCA TGTGCAATCC AAGTTTGGCA GACACCAAAA AGTGCATGAG 840

TTAGATAGAT GCTACTATAT TCTAATTCAG TGGTATTTAG ATTCAGTTGC ATAAATCGCT 900

TTTGTAAATC TGTACTAAGC ATGATATGAA GTTTATTTCG TAAGAAATTT TGGATTTCTT 960

TAGTCCCATT TTCAGAAAGA AGGGCAGCCA GAAGTGGTTC TGACTCTAGA TATTCAAAAA 1020

CTTCTAAAAT AGCGTCTCTT TTGTGATGAG CATGTTTTTG AAAAATATAT TCAAATGTAT 1080

GGAATAGCTT GCTTTGATAG TGCTCAATCA TATCATACTT ATCCTTATAG TGAGTATAGA 1140

AGCTGGAACG ACTAATTCCG GCTTTTTCTA CTAATTTGAC AGTAGAAATT TTATCAAATG 1200

GCTGTTCCAT CAGTAATTGT ACCATAGCAT TTTCAATAGT TCGCTTTGTT TTTAAGCGTT 1260

TGTTACTTTC TTGCATATTT CCTCCTTGTA AACAAATTAG ACTATATGTC TAAAAATAGA 1320

TTTTTTATCT TGTAATTTAG ATTTTTTAAT GTATAATCTA TTATATCAAA ATTTTAGACA 1380

ATATGTTTAA AAAAGGAGAA ACTAAGTTTA AAGAATGGAA AGCAATTTAA AAAAAACCAA 1440

CCTTTATTAT TGTCATGATC GGGATTTCTC TTATTCCAGA TCTGTACAAT ATCATATTTT 1500

TGTCATCAAT GTGGGATCCA TATGGGCAAT TGTCTGACTT ACCTGTGGCA GTTGTAAATA 1560

ATGATAAAGA GGCTTCCTAT AATGGTAATA CTATGGCAAT AGGAAAAGAC ATGGTGTCCA 1620

ATTTAAAAGA AAATAAAACC TTGGATTTTC ATTTTGTAGA TGAAGAGGAA GGAAAGAAGG 1680

GATTGGAAGA TGGCGATTAC TATATGGTAG TGACTTTACC AAGTGATTTA TCTGAAAAAA 1740

CAACTACATT ATCCAATATT CAATCGACAG CAGCTTATCA ATCATTGACA AGTGAGCAAC 1800

AAACTGAGAT AAGTGATTCT GTATCTCAAA ATTCAACTGA TAGTATTCAA TCGGCTCAGT 1860

CAATTGTAGC TTTAGTACAA GATTTACAGG GAAGTTTAGA AAACTTACAA AATCAATCTT 1920

CTAATCTTTC GACTTTAAAA AATCAATCTA ATCAAGTATC ACCTATTACT TCTACTTCTT 1980

TGATAGGATT GTCAAGTGGA TTAACAGAGA TACAAGGAGA TGTTACTAGC AAATTAGTTC 2040

CTGCCAGTCA GTCGATTGCA TCAGGTGTAA ACGCATATAC TACAGGTGTT GATAAAGTTT 2100

CTCAGGGCGC AAGTCAACTA AGTGAAAAAA ATGCCACCTT GACAGGTAGT TTGGATAAAC 2160

TAGTTTCAGG CTCAAACACC TTGACACAAA AATCTTCTAG ATTGACAGCA GGAGTTGGTT 2220

AATTACAATC AGGATCTGGG CAATTAGCAG ACAAATCCAG TCAGTTACTT TCAGGTGCTT 2280

CTCCATTAGA GAATAGAGCT AATAAATTGG CAGATGGATC TGGGAAACTA GCAGAAGGTG 2340

GAACAAAGTT AACTTCTGGA TTGGAAGATT TACAGACAGG ACTTGCTTCT TTAGGACAAG 2400

GACTAGGTAA TGCTAGTGAT CAACTCAAAT CAGTATCAAC AGAATCTAAA AATGCAGAGA 2460

TTTTGTCAAA TCCACTCAAT CTTTCAAAAA CAGACAATGA TCAAGTTCCT GTAAATGGAA 2520

TCGCAATAGC TCCTTATATG ATATCAGTTG CTCTTTTTTT GCAGCAATAT CAACAAATAT 2580

GATATTTGCG AAATTGCCTT CAGGACGTCA TCCAGAGAGC CGTTGGGCTT GGTTGAAATC 2640

TTGAGCTGAA ATAAATGGTA TTATAGCTGT TTTGGCAGGA ATTTTGGTAT ATGGAGGAGT 2700

TCAGCTTATT GGTTTAACTG CTAATCATGA GATGAGAATA TTTATTCTCA TCATCCTAAC 2760

AAGTTTAGTA TTCATGTCTA TGGTGACCAC TTTAGCAACG TGGAATAGCC GTATAGGAGC 2820

TTTTTTCTCA CTTATTTTGC TTTTACTACA GTTAGCATCA AGTGCAGGTA CTTATCCACT 2880

TGCTTTGACA AATGATTTCT TTAGATCTAT TAATCCCTGG TTACCAATGA GCTATTCAGT 2940

TTCGGGATTA CGACAAACAA TCTCTATCAA CAAGTCATTT TCCTAGCTGT CATACTAGTT 3000

CTATTTACTA GTTTAGGTAT GCTAGCCTAT CAACATAAGA AAATGGAAGA AGATTAAAAA 3060

AATCGACCGA TTAACTGGTC GATTTTTTAT GCCTTAGATG ACTTTCGTCT GTGATTATAG 3120

ATTCCAAATA GTAAGAGAGA AGTAAAGGAA CAGATTGCTC CAGTAATAAA ACCATTGGGA 3180

[illegible] TGGAGTT 3240

TGAGCTTGTT TAATTTCTAT TTTCTTACCA TCTTGGTAGG CAGACCAACC TTTGTCATAA 3300

GGAATGGTGA AGAAAATAGA TGTATCTTGT TGGACATCAT ATGTAGCAAA AACCTTGTTT 3360

TTAGAAGTTG ATACTGTGAC AGGTTGTTCT TTAATTTTTT GAATTGCCTC GGTGAAAGTT 3420

TTGGTATCTA AACGATAGAA GGTAGGAGAT TCAAATGATA CTTGTGAATT TCCAGGGAAA 3480

CTAACATTGA TATTGAAAGT TTTTTTCTCT TTAGTATATC CTAGATTAAA GAAGGAGAAG 3540

ACATTATCAG TTGTAAAAGT CTTTTTTTCA CCATTTACAA GGATGTCAAC CTTCTTTTGT 3600

TTATCGTTAG AAAAGTGAAG GTTTATGAAA GAGAGATAAA CTTGGCTGTT TTCTGGAACT 3660

TCAATTTGAT ACTGGATTGC TGCATCTTCA TTTGAAGAAC TTGTGACACT AATCAAATCA 3720

TTAGTATTTT CTATTTTTTC TGTTTTTTCA TAAGGTATTG GAGAAAAATA ATCAAAATTG 3780

ACGTTAGCAA GTTGATTTAA AAATGAGGCC TGATTATCCA AGGTATGTTC ATTGAACTTG 3840

ACATCATTGT AAACAGATTG ACTCGCAACT GCAATCGGAA GAGAGTATTG ATTTTCATAT 3900

AGGGTAAGAT TATCTTTTTG ATAGATATCT TTAAAGCCAT ACTTATCAAT AGGACTGTCT 3960

GAGATATTGT ACTGGATACC AAATAAACTA TCAGCCAAAA TACTATTATT TGCATATCGG 4020

AGATTGAGAT TAGTCCCAGA GGATTTAAAA CCAAGTTTAT CTAAAGTAGA GCTTGATGAA 4080

CGATTTCGAA CAGATGAAAA TTGAGAGATT CCATTGTAGT TGAATTTCAT ACTGTCATTT 4140

CCTGTCTGAG TTTGTAGTTT TTCAGTACGA GTAAATTGAT TTCCAATATA TGTTGAGAAA 4200

GATTCCATAG CTGGGATATC TCGACTATAA GCACTTCGAG AAGCAAATCC CCATTCCTTA 4260

GCAATTCCGT CCATTTGAGA TGAAGCATTT AAACTCATTT CAACCAGTAT AAATAAAGAG 4320

ATTAGAATGG CAAATAGATT CACAGATATA AACTTTTTGA TAACTGCAAG GAGTAAAAGA 4380

GAATAGACAA CCAAAAATTC AAGAGTAAGC AGAATATTCA AATCTGTTAA AAAAGAATAA 4440

TGCGATTTTA GATAGATGGT AGCTAAAAAT CCTGCTACTA CAAGAAAAAG CGAAACTAAA 4500

AAATTCCAGA CTTTAAGTTC TTTCAGACGC TTTAAGACTT CTGCTGCTGT GTAAATTAAC 4560

AAGGTAGAGA AAATCCAAGC ATAGCGATGT AAAAACATGT TTGGAGTATG CATGCCTTGC 4620

CAAAATAAGT CAAGAGCTTC TATGTAAAAG CTTGCAATTA GAAATGCAAA GAATATTACA 4680

TATATGAGTT TCACGTGAAA CTTAATAGAT TTCAGCGTAA AAAATAAAAT GGTCAAAATA 4740

AAGGGAAATA GTCCAACAAA AATCATTGGG ATGGCCCCAT ACTTTGTTGT GTCAAAGGAA 4800

CCAATGAATT GCTTAGCAAA GAGATCAAGA TACCAGCTAC TTTCAGTTTG AAACTTTGTA 4860

ACTTCAGTCA ATTTTTCCCC ATGTGTCTGT AAATCAAATA GAGTGGGAAG AGTCATAATC 4920

AAACTAGCCA TACCAGCTAA AAAGGAGATA ACTATGAAAT CAAGAACAGA TGATTTTCGA 4980

GTCTTAAAGT CCCACGAAAT TTGACAGAGA TACCAGAAAA TAAGAAACAA TACTGTCATA 5040

TATCCAAAAT AATAATTTTG AATAAATAAG ATTGACAGAC TTGTAAAGTA CAATAGGAGT 5100

TTCTTTTCAG TTATCAGTAG ATGTAAACCA GTTATAATTA AAGGAATCAA GATAAAAACA 5160

TCTAGCCAGG TTTTTATCTC TAATTGACTG ACAGTGAAAC TCATCAGAGC ATAGGAAGTA 5220

GATAAGGCTA GTTTTAAAAT CTGAGGGATA GATTGAAACA ATTTATTCAA ACTAAAAAAG 5280

GTTGACAGAC CAATCAATCC AAATTTTAAG AGAGTTGTCA GATAGATAGC ATCTGGCATA 5340

TTCGTTAGAT CAAAAAAGTA AACCAGAGGC GCGAGAAAAC TACCCAAGTA ATAACTAGAT 5400

AGGGCATAGA AGTTTAGCCC TAGACCACTT GTAAAGGTGT AAAACAGATT ACTATTTCCA 5460

TGTAGGATAT TTCGTAAGGC TACATCAAAA ATAACGTATT GATGAAAGCC ATCTCCTAAT 5520

AGAGGAGAGT TGTCGCTATT CCAGTAGATA CTTTGAGATA GATATACTCC AGACATAATC 5580

ACTACAGGAA TGATGAAAGA AATAAAATAG GTTCGATATG TTTTTAAAAA TGATTTCATG 5640

TTACCTCGTA GAATGATAGA AAACTCAGTT GGTTAACCCA ACTGAGTTTT GAAGTTTTAT 5700

TTAGTCTTTC CAAAGTTCTT TAACTTTTGC TTGTACTTCT GCATTTTCTA GGAATTCATC 5760

GTAGGTTTCA TCGATACGGT CAATGACGCC ATTTTTAGAT AAGACAATGA TATGGTTAGC 5820

CAAAGTTTGA ATAAATTCGT GGTCATGGCT GGCAAAGATG ATTGATTCTT TAAAGTTTTT 5880

CAATCCATCA TTCAAGCTTG AGATAGATTC CAAGTCCAAG TGATTTGTTG GATCATCAAG 5940

TACAAGGACA TTTGATTTTA AGAGCATGAG TTTTGAAAGC ATGACACGAA CTTTTTCTCC 6000

CCCTGACAAG ACATTTACAG GTTTGTTAAC TTCATCTCCA GAGAAGAGCA TACGGCCGAG 6060

GAAGCCACGT AGGAAAGTAT TGTCATCTTC TTCTTTACTT GCGAATTGAC GCAACCAGTC 6120

AAGAATTGAT TCTCCTCCTG CAAAATCAGC TGAGTTATCT TTTGGTAGGT AAGATTGACT 6180

AGTTGTAACT CCCCACTTGA CAGTTCCTTC ATAGTCAATA TCTCCCATGA TTGCACGAAT 6240

TAATGCAGTC GTTTGAATAT CATTTTGTCC AATAAGTGCT GTCTTATCAT CTGGACGCAA 6300

GATGAAACTA ATATTATCCA AGATAGTTTC ACCATCAATC TTTACAGTTA AATTTTCTAC 6360

TGTCAAGAGA TCATTACCAA TCTCACGTTC CGCTTTAAAG TTGATAAATG GATATTTACG 6420

ACTAGATGGC ACAATCTCTT CTAGCTCAAT CTTATCAAGC ATTCTCTTAC GTGATGTTGC 6480

CTGCCTTGAC TTAGAAGCAT TGGCAGAGAA ACGAGCAACA AATTCTTGCA ATTGTTTAAT 6540

TTTTTCTTCT GCTTTAGCAT TACGGTCTGC TAGCAATTTA GCAGCAAGCT CAGAAGATTC 6600

CTTCCAGAAG TCGTAGTTTC CGACATAGAG TTTGATTTTT CCAAAGTCAA GGTCGGCCAT 6660

GTGAGTACAA ACTTTGTTTA AGAAGTGACG GTCGTGGGAT ACTACGATAA CTGTGTTATC 6720

AAAGTCAATC AAGAA[illegible] 6780

CTCGTCCAAG AGAAGAACAT CTGGTTTACC AAAAAGTGCT TTGGCGAGGA GAACCTTTAC 6840

TTTTTCACCG TTGGCCAATT CGCTCATGTT TTGGTAGTGT AATTCTTCTG GAATGTTTAG 6900

GTTTTGAAGT AGTTGAGAGG CTTCACTCTC TGCTTCCCAA CCTCCAAGTT CGGCAAACTC 6960

TCCTTCGAGT TCGGCAGCAC GAACCCCGTC CTCGTCTGAG AAATCTTCCT TCATGTAGAT 7020

AGCATCTTTC TCTTTCATGA TGCTATAAAG TTTTTCATTT CCCATGATAA CGACATCAAT 7080

GGCACGTTCA TCTTCGTAGT CAAAGTGATT TTGACGAAGA ACAGAGAGAC GTTCATCTGG 7140

ACCAAGAGAG ATGTGACCAG TAGTAGGTTC GATATCTCCA GCTAAAATTT TTAAAAAGGT 7200

TGATTTTCCG GCACCATTAG CACCGATTAA TCCGTAAGTA TTTCCTTCTG TAAATTTGAT 7260

ATTGACATCA TCAAAAAGTT TGCGATCACT AAAACGTAGT GAAACATCAG ATACTGTAAG 7320

CAATGTTTTT CTCCTATATG TGTAATATAT TTATTCTACT AGAAAATACA GAAATATTCA 7380

AATTTTTATT TGTCAATTTT GTGTAAATTA TATTTACAGT ATCCTTTACA CAAATCTGTA 7440

AAAAGCAAGG CTGATTTATT TTGATAAATT ACGGTTATTT CATTAAAAAA ATGCTATAAT 7500

TGAAAGGACT ATATCGAAGG AGAACAAAAT GACTAAACCC ATTATTTTAA CAGGAGACCG 7560

TCCAACAGGA AAATTGCATA TTGGACATTA TGTTGGAAGT CTCAAAAATC GAGTATTATT 7620

ACAGGAAGAG GATAAGTATG ATATGTTTGT GTTCTTGGCT GACCAACAAG CCTTGACAGA 7680

TCATGCCAAA GATCCTCAAA CCATTGTAGA GTCTATCGGA AATGTGGCTT TGGATTATCT 7740

TGCAGTTGGA TTGGATCCAA ATAAGTCAAC TATTTTTATT CAAAGCCAGA TTCCAGAGTT 7800

GGCTGAGTTG TCTATGTATT ATATGAATCT AGTTTCGTTA GCACGTTTGG AGCGAAATCC 7860

AACAGTCAAG ACAGAGATTT CTCAGAAAGG ATTTGGAGAA AGCATTCCGA CAGGATTCTT 7920

GGTCTATCCA ATCGCTCAAG CAGCTGATAT CACAGCTTTC AAGGCTAATT ATGTTCCTGT 7980

TGGGACAGAT CAGAAACCAA TGATTGAGCA AACTCGTGAA ATTGTTCGTT CTTTTAACAA 8040

TGCATATAAC TGTGATGTCT TGGTAGAGCC GGAAGGTATT TATCCAGAAA ATGAGAGAGC 8100

AGGGCGTTTG CCTGGTTTAG ATGGAAATGC TAAAATGTCT AAATCACTAA ATAATGGTAT 8160

TTATTTAGCT GATGATGCGG ATACTTTGCG TAAAAAAGTA ATGAGTATGT ATACAGATCC 8220

AGATCATATC CGCGTTGAGG ATCCAGGTAA GATTGAGGGA AATATGGTTT TCCATTATCT 8280

AGATGTTTTT GGTCGTCCAG AAGATGCTCA AGAAATTGCT GATATGAAAG AACGTTATCA 8340

ACGAGGTGGT CTTGGTGATG TGAAGACCAA GCGTTATCTA CTTGAAATAT TAGAACGTGA 8400

ACTGGGTCCG G 8411

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9064 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

TGCCGTACTC AAGTACAGCC TGCGCTAAGT TTCCTAGTTT GCTCTTTGAT TTTCATTGAG 60

TATTAGTAAC CAAAATCCGA CCACATAGCC AGCCCCTATG AATATAGCCA TTAAAGCTAG 120

CATGGAATTT AGGAAATTAA AAACCACCGC AGATACAAAG GTTAGCACAA AAACATTAAA 180

AGCAATGGTG TCAGAAGCCA AGACTAGAAT ATAGGGTGTC AACCGATCTA AAGTTTTGGA 240

ATCTAGGAAA AATAAGTGTT TATACATGAT GACCTCCTCT ATGGCTGAAA AGCAAGCCTT 300

TTGTTTTTTT ACCCCAAGAC CCTATGTAGA AAAGTGAGCA AAAACGGGAA GGTCGCTACA 360

ATATTATTGA TCACATGCAC CGCATAGGAT GGATAAATGC TCTTGGTATA GCGGGTCAAA 420

CCAGCAAAGA TGATTCCAAC TGTTGCAAAG ACGAAGATAT CTAACAGACT AGGCAGGCTT 480

GAAAAATGAG GGAGAGCAAA TAAAATAGAA GGAAGAAGCA AATCAAGACC AAATCGCGAA 540

TGCTTAAAGA AAGCATGTTG CAGTAATCCT CTATAAATCA ATTCTTCCAT CAGTGGAACC 600

AGAAAGAACA GGGCTATATA AATACCTAGC TCTGCAAAGT TAGTCCCACT ATAACCAATC 660

AATACAGCCC AACCTTCCGC AGTTGACTGA ACATGTTTAG CTGTCTGAAC GTTAAAAGAG 720

ATCTGGAACA CTAGCACTAA TACTGTCAAA ATCGAATACC AAAGCCATTT TTTTCTTGGA 780

ATGCGGAAGA GATAACCATG GCCTGTCTTA ACAAGAACCA CAATCATGAC TCCAATAAAA 840

AGTAAACTCA AGATATTTTG AATCCAGAAT AAATTGCCTA TCTGAGAAGA AAATTGCCAA 900

TAGTTTTGGA CGATAAGCGT CAGCTGAGAA AGACTAAATA CGAAAAATAA GTAAGAGAAG 960

ACTGCACTTA TTTTGAATAG AAGTTGATAC TTTTTCATAG AAATCCTCCC TACTATGACC 1020

TCACCTTGTC AGGCTCTACT GCTGTAAGAT TAAGAAGACA GTTTGTTTTT TTTAAGGCTA 1080

ACCTGACTAC TAGATAATAG ATACATTAAG GCATTAAAGA CAATGAAAAT ATGTCCATAG 1140

AATAAAATCA ACCTCGCATC CAAACCAAGA TAAAGTTTGA TTATCAAAAA GATGAGCAAA 1200

AGAATTTGAA ACCATAAGGT TTTTCCAAAA ATAAATTTAA AGCGATTTCG AATATCTACT 1260

TCCTTGATTT TTACCGCCAC CCCTTTATTA GCAAGAAGGA AAACTCCTGC TTCAAACAAA 1320

CCACTGTAAA GAACAAGCCA CCCAATAGAT ACGATAGAGA TTTGTAAAAA TGTCCCTAAA 1380

AGAATATCCA ACACACTACT CAAGAAAATA ACAAAAAATA ATCTGTATTT CATATTAAAT 1440

ACCTCCATTC ATTTATTTCA CTAACAATTT AATAGAGCCT TCTACTCAAA TATCCTGTCA 1500

GAAAAGGATA GAAAGCTACT TTTTATAATA CTTCAAGCCC CACATGAGCA GAAGCGTGAT 1560

AAACAAGCAG AGAATACACC TATATAAGCG ATTAGTTGTT GATAGAATTC TGTTTCTGAA 1620

ATACCTCTAT ACAAACAAAT GACAAACATA AAATCTGCCA AGCCGATAAA CATAAGTTGA 1680

TTGGTTCTAG GACTAACCAA ATCATCATTT ACTTATATTT AAGAGTATCT CTTTTATTTT 1740

AATGTATGTT AGCACTGAAA AGCAAGACAG GCCAATAATA TTTAAAATGA ACAGTAACGG 1800

GGTTAAGTCT CTAAAAAAAT TATCTACTGA CACTACAAGA AATACTATAC ATATTATAGT 1860

CGAAACTATC TTTTTCTTAT CCATAATTAT TTACTCCTTT CCTAACAAAT CCAGCTTATC 1920

AATCAAGAGC GATTTTTAAC ATAATGTAGC AGCACCCGTT GCAACTTTGA CAAGTTTAGT 1980

ATATCATTGT TTTTTAAAAT TTTTCATCCA AATCTTGAAT TGTCATCGAA ACATCTTGAA 2040

TTGTTAAAAA ATTTAAAAAG TAAGCATTAA AAACATACTT TCCTCTTTAT ATTGTATTGA 2100

TACCAACTTG TTTGTAGACT TTTCATCCTG CTATCACATA TCATTTTGAC AGGCGAAACA 2160

ATATTAAAGA AACTCCCCTG TAAATTAAGC TAGCAAATAC AGGGGAGAAA TTTATTTTTT 2220

AGAGAGTACT ATCCGTATCC TTTTTGGAAG ATTTTGAAAA TATTTTTCTA ATTAAGTCAT 2280

CCATATAAGG ACCAAATATA CCAACTACTA AACCAATAAT AAAACTTTTA AAATCCATAA 2340

TTACCACCAA CATATTGCTG CATAGGCTAC ACCTCCAAGT ATAGCTCCAC CTGCAGCACC 2400

AGTTACACCT ATTCCTATAG CAAATGGTCC CAATAGAAAT GTCAAACCGT TGTTGCACAC 2460

CCATCAATTG CGCCATATGC AACCCCTGCT GCACAACTAA TTTTTCTTCC CCAATCAATA 2520

TCTCCACCTT CAACGCAAGC AAGCATTTCA TTATCCATAA CTGCAAATTG TGACATCATT 2580

TTTGTATCCA TATAGTGTAT CACTTTTCAG TTACGGAACA AGTTTAATAT AAAAATTATC 2640

AAAAAAACAT AGGCAATAAA GAGAAAAATT AATTTATCAT AGATTAGAAA TAATATGACA 2700

AAACAATTCA ATGATGTTAA TTCAATAGTC TTTTGTTTTT TATCGGAGAT ACTTATGGAT 2760

AGATAAATAA GATAGGTTTG AAAAGCGAAG AGAATAATAA AGAATATAGC CTTCATAAAA 2820

TTTAGCTTTC ATTTTTATGA TGTAGCGGTA TAGGCTAAAT ATCCACAAAC CACTGCTCCT 2880

CCAATTCCTC CTATTGCAGC GCCCCATGGT CCTAGAAGTC TCCCATATTT CACTCCACCC 2940

GCTGCACAAC CTAAAGCAGC AACTACAGCT GCTCCTCCGG AATTACCTCC ATAAACCTCA 3000

CTCAGCATTG TTTCATTTAT ATTACAATAA GTATTCATAC AAGTCTCCTT TTATTAAAAT 3060

CCACCCGTTG CCCCTGTTAC TCCTGCCCAA AGATCCACAC CAAATTTAGC TCCTATGTAT 3120

CCACATGCTC CCATAAATGG TGCTCCAACA CCACTCGCAG CACAAATAGC TGTCCCTAGC 3180

CCCCAGCCAC CAAAAGCAGC ACCACCACCT TCTAAGACAT TAGTTTGCCA ATTATTCTTG 3240

CCTCCTTCAA TACTAGATAA CATAGTTATA TCCATTTCAT GAAATTGTTC CATAATTTTT 3300

GTATCCATGA CAAATACTCT TTTTTATTTT TAATTTTTGT CTTGTTGTAA CTTTGACAAG 3360

TTTAGTATAT CATCGTTTTT TAAAATTTTT CATCCAGATT TTGAATAGTC ATCGAAACGT 3420

CTTGAATTGC AAAAATTACA TTAGACTTCC TGCAAAACTA GAATCCTAGT TCATGATTGA 3480

TAATACCAGC ACTCAAATTC ATTCGTAATC CGAAGCGTTT ACGATGACTT CGATAGGTTG 3540

TTGAAAACAT TTTAAACGTT TTTACTTTGG CAAAGATGTT CTCAACCTTG CTTCTCTCCT 3600

TAGATAGCGC ATGGTTACAG GCTTTATCTT CAACTGTTAG CGGTTTGAGT TTGCTGGATT 3660

TACGTGAAGT TTGTGCTTGA GGATATATCT TCATGAGCCC TTGATAACCA CTGTCAGCCA 3720

AGATTTTACC AGCTTGTCCG ATATTTCTGC GACTCATTTT GAACAACTTC ATATCATGAC 3780

AATAGTTCAC AGTGATATCC AAAGAAACAA TTCTCCCTTG ACTTGTGACA ATCGCTTGAG 3840

TCTTCATAGC GTGAAATTTC TTTTTACCAG AATCATTCGC TAATTCTTTT TTTAGGGCGA 3900

TTGATTTTTA CTTCCGTCGC ATCAATCATT ACCGTGTCCT CAGAACTGAG AGGAGTTCTT 3960

GAAATCGTAA CACCACTTTG AACAAGAGTT ACTTCAACCC ATTGGCTCCG ACGGAGTAAG 4020

TTGCTTTCGT GAACACCAAA ATCAGCCGCA ATTTCTTCAT AAGTGCGGTA TTCTCGCACA 4080

TATTGAAGAG TGGCCATAAG AAGGTCTTCT AGGCTTAATT TAGGTTTTCG TCCACCTTTT 4140

GCGTGTTTAA GTTGATAAGC TGTTTTTAAT ACAGCTAGCA TCTCTTCAAA AGTCGTGCGC 4200

TGAACACCAA CAAGACGCTT AAATCGTGCA TCAGTTAGTT GTTTACTTGC TTCATAATTC 4260

ATAGAACTAT AGTAAAATGA AATAAGAACA GGATAAATCG ATCAGGACAG TCAAATCGAT 4320

TTCTAACAAT GTTTTAGAAG TAGAGGCGTA CTATTCTAGT TTCAATCTAC TATACTATAC 4380

CATATTTTGT TTCGCAGGGA ATCTATTATA AAAGGGTAAG TATTGCAAAA ACACTTACCC 4440

TTTTCTTTTA TACTTCATTA AGCTCTACTT TTTATAATAC TTCAAGCCCC ACATGAGCAG 4500

AAGCATGATG ATTAAGCAGA GAACAGCGCC AATATAAGCG ATTATTTGTT GGTAGGATTC 4560

TCCTGCTGTG ATACCTCTAT ACAAACAAAT AATAGACATA AAACCTGTCA AGCCGATGAA 4620

CATAAGTTGA TTGGTTCTAG GACTAACCAA ATCATCATCT TCAAACTCTC TTATCCTCAT 4680

TTCCCTAGTG AGATAAACAG TAACCAAAAT AGAAGCCAAG TTAATAACTA CTAAAAGAAA 4740

TTGGAAAACT ACGGAAAAAT TTAAAAACTG ACGAGATAGA AATAGATAAG TAGAAACAAG 4800

CAAGGGCAAC TGACCTAAGA ACAATCTCGC AAGGAAGATG TTCCGTTTTT TAGCAAGAAA 4860

AGTTTTCATT TCTTTTCTCC TTTCTTTTTA TTGATAGCAA AATAGATCAT AACTGCAATC 4920

ACATAGGCTA TGGTATAAAA TAGCTGATAC CAAGCACTCT CCCTAAGCGG ATATAGAAAG 4980

ATGGACATGA TTAGATACAG AACGAAAATA ATCAGTATTT TTTTCTTCAT AAGATTTCCT 5040

CCTAAATGTG CGATTTATCT TAGTTGAGCA AGAACATTTA CACTGCTAGT ATAGCACTTA 5100

TTTTGACCTT GGATCACTCA AATCATAAAT GGTCATCAAA ACCTCTTGAA TTGTAAAAAT 5160

TAAAAAAGCA AGCATGAAAA ACATACTTTC CTCTTTATAT TGTATTGATA CCAACTTGTT 5220

TGTAGACTTT TCATCCTGCT ATCACATATC ATTTTGACAG GCGAAACAAT ATTAAAGAAA 5280

CTCCCCTGTA AATTAAGCTA GCAAATACAG GGGAGAAATT TATTTTTTAG AGAGTACTAT 5340

CCGTATCCTT TTTGGAAGAT TTTGAAAATA TTTTTCTAAT TAAGTCATCC ATATAAGGAC 5400

CAAATATACC AACTACTAAA CCAATAATAA AACTTTTAAA ATCCATAATT ACCACCAACA 5460

TGTTGCTGCA TAGGCTACAC CTCCAAGTAT AGCTCCACCC GCAGCACCAG TTGCTGCACC 5520

TTGCCATGTT CCTGTTTTAA TGCCTAGTTG AAGACCTCTT GCTGCTCCTC CTCCAACACC 5580

TGCTTTGGCA AAATCTCCCC AATTGCATCC GCCACCTTCA ACGCAAGCAA GCATTTCAGT 5640

ATCCATAACA GAAAATTGTG ACATCATTTT TGTATCCATG ACAAATACTC CTTTTTTAAA 5700

AAACTAAAAT AAATCAGAAT AGAATCCTCA TAATTTTACT ATAAGTCTTA CCAACTTAGT 5760

CCCAATTTAT CACCAACCAT ACCTCCTAAG CATGTTAATC CACCCCCAAT TGCACCAATG 5820

TGTGCTCCAA CAAATGCACC AGCAAGTCCA GCTACTCCTA AAGTGGCCAA ACCTGCTCCA 5880

GTTCCACCAG TTATAATTCC CGTAGTGACT CCTGTAATCA GTGCATTTTG ACAATCAGTG 5940

GAGCTATACC CCCCTTCAAC TTTCGCAAGC ATTTCAGTAT CCATAACCTC TAACTGTGAC 6000

AACATTTTTG TATTCATGAT GAATACCTCC TTTTTATTTT CAATTTGTTA CCAAAGTCTT 6060

AAATTCAATA AACAAATAGA TTTTTTATAG TATCTTTTTG ATTTTCTTAA AAAAGTATAT 6120

ACGTCTACTA TCTTCTTAAA GGTAGCAGTA CCTATTTTTT AGTCTAAGAT TTCAATAATC 6180

TTGAGTATCT AAAATATCTT AATTTCGTTA TTCTCCTTGC AATAAAAAGT TTTACTATAC 6240

TATTTATTAA CTTGCAGAAA GCAAAAAATA TTAGTAAATA ATAGTTTATA GTTAAGTTTT 6300

TTATTCCTAC CAATCCATCA ACTAAGTAAA GCATCAACGA TTACATAAAC GATTGATAAT 6360

ATAATTAAAA TTTTGCTAAC TATCTTATTC TCATCATTCT TAGATAACTT TGATATTTTG 6420

TAAGTAAGTA AATAAGACAG TAAATTAATA GCGATAATAA TACTATATTT AAGAATCATA 6480

ATCTTACAAA GAGGACATAA TTCCTGAACC TACACAAATA AGTGTTGCTG CTCCCCCAGT 6540

TATCGGACCA GTCGCAGCAG CTAATAGTAC TGCTCCAATA CAACCACCGA TTGCAGATCC 6600

TAAATTGCCT CTTCCTCCAC TAACTATTTC GAGTTCTTCA TTATCCATAA CAGAAAATTG 6660

TTCCATCATT TTTGTATTCA TGACAAATAC TCCTTTTTTC TTTTTTTATT TTTGTCTTGT 6720

TGTAACTTTG ATAAGTTTAG TATATCATCG TTTTTTAAAA TTTTTCATCC AGATCTTGAA 6780

TTGTCATCGA AACGTCTTGA ATTAGCTTTT TTATTTCAAG CCACCTCTAA ATGTTTAAAA 6840

AAAATAATTT CTAATCACTT TTTTACCATT CAGGAAGTTT TAATGACTAT TCAAGATTTC 6900

ATAAAATATG AACTTAGTTT TATGACATAA TAGACCTATC CACTATATGA AAGGAATTGC 6960

CAATGACTTC TTATAAACGT ACATTTGTTC CTCAAATAGA TGCGAGAGAC TGTGGTGTCG 7020

CTGCCTTAGC CTCGATTGCT AAATTCTATG GTTCAGATTT TTCTCTAGCT CACTTGAGAG 7080

AACTTGCAAA GACCAATAAA GAAGGGACGA CTGCTCTTGG CATTGTAAAA GCCGCTGATG 7140

AAATGGGCTT TGAAACAAGA CCTGTTCAAG CAGATAAAAC GCTCTTTGAC ATGAGTGATG 7200

TCCCCTATCC ATTTATCGTT CACGTTAACA AAGAAGGAAA ACTCCAACAT TACTATGTTG 7260

TCTATCAAAC AAAGAAAGAC TATCTGATTA TTGGTGATCC TGACCCTTCT GTAAAAATCA 7320

CTAAAATGTC AAAAGAACGC TTTTTCTATG AATGGACTGG AGTAGCTATT TTTCTAGCTA 7380

CCAAACCCAG CTATCAACCC CATAAAGATA AAAAGAATGG TCTACTAAGC AAGCTTCCTT 7440

CCTCTGATTT TCAAACAAAA ATCTCTCATT GCTTACATTG TTCTCTCAAG CTTATTGGTC 7500

ACTATTATCA ATATAGGTGG TTCTTACTAT CTCCAAGGAA TCTTGGATGA ATACATTCCA 7560

AATCAGATGA AATCAACTTT AGGAATCATC TCAGTTGGTC TGGTTATCAC CTATATCCTC 7620

CAACAAGTCA TGAGCTTCTC CAGAGATTAT CTCCTAACCG TTCTGAGTCA GAGATTAAGT 7680

ATTGATGTGA TTTTATCCTA TATTCGCCAT ATTTTTGAAC TTCCCATGTC TTTCTTTGCG 7740

ACACGTCGTA CAGGAGAAAT CATTTCACGA TTCACAGATG CTAACTCTAT TATAGATGCC 7800

TTGGCTTCTA CCATTCTTTC TCTTTTTCTG GATGTTTCTA TTCTGATTCT TGTAGGAGGC 7860

GTCTTACTGG CACAAAACCC TAATCTCTTC CTTCTTTCTC TTATTTCCAT TCCTATATAC 7920

ATGTTCATCA TCTTTTCTTT TATGAAACCT TTCGAAAAAA TGAACCATGA TGTCATGCAA 7980

AGTAATTCTA TGGTTAGCTC TGCCATTATC GAAGATATCA ACGGGATTGA AACTATAAAG 8040

TCGCTCACGA GTGAAGAAAA TCGCTATCAA AATATAGACA GCGAATTTGT AGATTATTTG 8100

GAAAAATCCT TTAAGCTCAG TAAATATTCT ATTTTACAAA CGAGTTTAAA GCAGGGAACA 8160

AAATTAGTTC TGAATATCCT TATCCTATGG TTTGGCGCTC AATTAGTCAT GTCAAGTAAA 8220

ATTTCTATCG GTCAGCTGAT TACCTTTAAC ACACTTTTTT CTTACTTTAC AACTCCTATG 8280

GAAAATATTA TCAACCTCCA AACCAAACTC CAATCTGCGA AGGTCGCTAA TAACCGTTTG 8340

AACGAAGTCT ATCTAGTCGA ATCTGAATTT CAAGTTCAAG AAAACCCTGT TCATTCACAT 8400

TTTTTGATGG GCGATATTGA ATTTGATGAC CTTTCTTATA AGTATGGTTT TGGATGAGAT 8460

ACCTTAACAG ATATTAATCT CACGATTAAA CAAGGAGATA AGGTTAGCCT AGTTGGAGTT 8520

AGTGGTTCTG GTAAAACAAC TTTAGCCAAA ATGATTGTCA ATTTCTTTGA ACCCTACAAA 8580

GGGCATATTT CCATCAATCA TCAGGATATT AAAAACATTG ATAAAAAAGT CTTGCGCCGT 8640

CATATTAATT ACCTACCCCA ACAAGCCTAT ATCTTTAATG GCTCTATTTT GGAAAACTTA 8700

ACCTTGGGCG GTAATCATAT GATTAGTCAA GAAGATATTC TAAAAGCTTG TGAAGTAGCT 8760

GAAATCCGTC AAGACATTGA AAGAATGCCT ATGGGCTATC AAACTCAGCT CTCTGATGGA 8820

GCTGGTCTAT CAGGAGGACA GAAGCAACGA ATCGCTCTCG CTCGTGCTCT TTTAACTAAA 8880

TCTCCTGTTT TAATACTAGA TGAAGCTACT AGCGGTCTTG ATGTCTTGAC TGAGAAAAAG 8940

GTTATAGATA ATCTTATGTC TCTAACTGAT AAAACCATTC TCTTTGTAGC CCATCGTCTC 9000

AGTATAGCCG AACGAACCAA CCGTGTCATT GTTCTTGACC AGGGGAAAAT CATTGAAGTT 9060

GGTA 9064

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7780 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

CTCCATTTTT TTGATTTCAT AAATAAACAA CCTCTCTGTT AATTTTGTAT AATTATAACG 60

ATATCCAAGT TACTTGTCAA GTGTTTTTTA AATTTTTATC TCAAAAATAT TTTTTCGTTC 120

AAAAAAAGGA GCCATCAGTT GATTTCAAGC TCCCTTTTAT ACAGAATTAA ACTATTTTAT 180

AGTTCGACAA TCTTACCTGT TTCAAAGTAG ACAACCCATT CACAGATATT TTTAGCATAG 240

TCACCGATAC GCTCCAAGTA GGAAATAACT TGGAAATAAT CACGACCCGT AACAATGGCT 300

TCTGGATTTT TCTTAATCTC TTCAGTCGCA AGGTCACGGA TAGTTTCAAA ATAGTGGTTA 360

ATTTGCTCAT CCATGGAGGC CACCCGGTAT GCGTCGTCAA CAGAACCATT AAGATAAAGA 420

TCAAGTGCTG CTTCCACAAC GCTTTTAACT TCACGTCCCA TTTTTTTAAT TTCTTCCTCT 480

ACAGCTGGAA TGCGCTCTTC CCCCTTCATA CGGATGGTTG CCTGGGCAAT GGCTACAGCG 540

TGATCCCCCA TACGCTCCAC ATCTGATACA GCCTTAAGGA CAGTCAAGAC TGTACGCAAA 600

TCTTGAGAGA CTGGTTGTTG GAGTGCGATC ATTTCAAATG ATTTCTTTTC CAGTTTCACT 660

TCGTATTCAT TTACTTCTGC ATCATCTTCG ATGACCTCTT TTGCCAGGTC ACGGTCATGC 720

GTGACAAAAG CACGTACCGT ACGATTGATT TGTGAGAGCA CTTCTTGTCC CATAGCGTAG 780

AACTGGTTAT GTAATTTCTC TAAATCTTCT TCAAATTGAG ATCGTAACAT CTTTCATCTC 840

CTTATCCAAA TTTTCCTGTA ATATAGTCTT CCGTTTCCTT GTGTTGGGGA TCAAGGAACA 900

TCTGCTTGGT ATCATTAAAT TCAATCAAAT CTCCATCTAG GAAAAATCCT GTCTTATCAG 960

AGATACGTGA AGCTTGCTGC ATGGAACGGG TTACCAGAAG CATGGTGTAC TTGTCTTTTA 1020

GACCATACAA GGTTTCCTCA ATTTTACCAG CTGAAATCGG ATCCAAAGCC GAAGTTGGCT 1080

CATCCAAGAG GATGATTTTA GGACTAGTTG CCAAGACACG GGCCACGCAG ACACGCTGCT 1140

GTTGACCACC TGACAATCCA ATAGCTGAAT CATATAGACG ATCCTTGACC TCATCCCAGA 1200

TAGAGGCACC TTGCAAGGCT TTTTCTACGG CTTCATCCAG AACCTGCTTA TCCTTAATTC 1260

CATTGATACG AAGCCCGTAG ACAACATTCT CATAGATAGT CATAGGGAAA GGATTAGGTT 1320

GTTGGAAAAC CATTCCGATT TCCTTACGTA ATTCAACCGT ATCTGTACGC GGACTGTAGA 1380

TGTTGTGACC ATTGTACACC ACGGATCCAG TTGTGGTCAC CTCTGGATTG AGATCTCCCA 1440

TGCGGTTGAG AGACTTGAGG AGGGTTGACT TCCCTGATCC AGATGGACCA ATCAAGGCTG 1500

TAATTTCCTT AGGTTGGAAA GATAGGGAAA CACTATTCAA AGCCTTCTTT TTATTATAAT 1560

AAACGGACAG GTCTGATACC TGTAAAATCG CATCTGTCAT ACGGTTTCCT TTCTAACCAA 1620

AGTGACCAGA TACATAGTCA TTGGTGGACT GTAGCTTGGC ATTTTGGAAA ATAGTTGCAG 1680

TCTTGTCATA CTCAATCAAA TCACCCAAGT AAAAGAAGCC TGTATAGTCA CTTGCACGAG 1740

CAGCCTGCTG CATATTATGC GTTACAATGA TGATGGTAAA GTTTTTCTTG AGCTCAAACA 1800

TGGTCTCTTC TAGTTGCATG GTCGCAATCG GATCCAAGGC TGAGGCTGGC TCATCCATTA 1860

AGAGGATATC TGGCTTAACA GAGATGGCAC GAGCGATACA GAGACGTTGT TGCTGACCAC 1920

CTGATAAGGT CAAGGCTGAC TTGTGGAGAT CGTCTTTAAC CTGATCCCAG AGGGCAGCCT 1980

GACGAAGGGA GGTTTCTACG ATTTCATCTA GGACTTGCTT ATCCTTAACT CCAGCACGTT 2040

CATGCGCAAA GGTAATATTA CGGTAAATTG ACTTAGCAAA TGGATTGGGA CGTTGAAAAA 2100

CCATTCCAAT GTGTTTACGC ATTTCATAAA CGTTGATTTC TGGACGGTTG ACATCAATTC 2160

CACGATAGAG AATCTGCCCA GTTACTTTAG CAATATCAAT AGTATCATTC ATGCGATTGA 2220

GACTGCGTAA GTAGGTAGAT TTCCCCGATC CCGACGGGCC AATCAAAGCT GTAATTTTAT 2280

TTCTTTCAAA TTGCATATCA ATCCCCTTAA TGGATTCATT TTTACCATAG TAAACATGGA 2340

CATCCTTAGT AGAAAGGGCT ACTTTTTCTT CAGGAAAGGT AAGGATATGC TTCTCATCCC 2400

AGTTATATGT TGACATGGCT TCTCCTTTAG GCAGCGGTTA ATTTCTTGTG TAGATAGCTT 2460

CCGAACTTAC GAGCTCCAAA GTTAAAAATC AGGATAAAGA TCAGGAGCAC AGCGGCAGAA 2520

CCTGCTGATA CAATGGTTCC ATCTGGAATA GTGCCTTCAC TATTGACTTT CCAGATATGG 2580

ACAGCCAAGG TTTCTGCTTG ACGGAAGATA GAGATGGGGC TAGTCACACT GAGGATATTC 2640

CAGTTAGACC AGTCAAGAGC TGGCGCCGAT TGCCCTGCTG TATAGATCAG AGCTGCAGCT 2700

TCGCCAAAGA TACGACCAGA TGCCAAGACG ACACCCGTTA CAATACCTGG AAGCGCTTCC 2760

GGAATAACAA CATGAACCAC TGTCTCCCAG CGAGAAATCC CAAGAGCCAG ACCAGCCTCA 2820

CGTTGGGTAT GGTGAACGTG TTTCAAACTA TCCTCTACAT TACGCGTCAT CTGAGGCAAG 2880

TTAAAGACTG TCAAGGCCAA GGCACCTGAA ATGATTGAAA ATCCATACTC AAACTGGACT 2940

ACAAAGATCA AGTAACCAAA GAGACCCACC ACCACTGATG GTAAAGAGGA CAAAATTTCA 3000

ATACAAGTCC GCACAAAGTT GGTAACAGGA CCTTTTTTAG CATATTCAGC CAAGTAAATC 3060

CCAGCTCCCA TAGAAAGAGG TACAGAAATA ATCAAGGTAA TGACCAATAG GAAAAAGGAA 3120

TTGTAAAGCT GAATGCCAAT CCCACCACCT GCTTGAAAAG CAGAAGACCT TCCAGTCAAG 3180

AAAGACCAAG AGATATGGGG CAAGCCCCGA ACCAAGATAT AGAGAATCAA GGAAGCCAAG 3240

ATTGTCACAA TGATGCTAGC AATCGTATAG AGGACAGCTG TTGCAAGTTT ATCTAATTTC 3300

TTAGCGCGCA TAATTTTTCT TTCCTCTTTC TTTCGTAATC AATTTAATCA CACTGTTAAA 3360

AACTAAGCTC ATCAAGAGCA GTACCAAGGC CAGTGACCAG AGAACATTAT TATTTACAGT 3420

TCCCATGACA GTGTTCCCAA TTCCCATAGT TAATATAGAA GTTAAAGTTG CAGCTGGTGT 3480

GGTCAAGGAA GTTGGGATAA CAGCTGAGTT TCCGACAACC ATCTGGATAG CTAGAGCCTC 3540

ACCAAAGGCA CGCGCCATCC CAAAGACCAC TGCAGTGAAA ATACCAGAAC GGGCCGCCTT 3600

CAAGATCACA CGCCAGATAG TCTGCCAGCG AGTGGCTCCC ATAGCGAAAC TGGCTTCACG 3660

ATAATAACGA GGAACCGCAC GCAAGCTATC CGTTGTCATA AAGGTTACGG TCGGCAAAAT 3720

CATGACAAAG AGGACGGAAA TCCCTGACAA AATCCCAAAA CCAGTCCCAC CAAAGACACT 3780

GCGAACAAAG GGAACGACGA CTTGCAAGCC AATAAATCCG TACACTACTG AAGGAATCCC 3840

AACCAGGAGT TCAATAGCTG GTTGCAAAAT CTTCGCCCCT TTTGGTGATA CTTCGGTCAT 3900

AAAAACTGCT GCACCAATAG CAAAGGGTGT TGCGATAAGG GCTGAGAGAA TGGTAACGAT 3960

AAAGGAACCC AAAATCATAG GAAGGGCACC AAATTCTTTA CTAGAAGGAT TCCAAGTTCC 4020

TCCCAAAAGA AAGTCAAAGA TATTCACACC ATTGACAAAG AAGGTCGACA AGCCTTTTTG 4080

CGCTACGAAA ACCAAAATCA TGGCCACAAG GATGACTATC AAAGAAAGAC AGGCAAAGGT 4140

CAAACCTTTT CCTAATTTCT CCAGACGAGA ATTCTTTGAT GGAAGCAACA TTTTCTTAGC 4200

TAATTCTTCT TGATTCATTA TTGTCTCCCT TCCAACACTG TCACAGTTCC GGCAGCATCT 4260

TTTTCAACCT TCATTTCCTT AATCGGAATA TACTTCAATC CTTTGACAAT CCCTTCTTGG 4320

GTCTCATCCG AGAGAACAAA ATTGAGAAAT TCTGCAGCCA ACTCATTGGG CTGCCCCAAT 4380

GTATACATAT GCTCATAAGA CCACAAGGGC CAATTATTGC TACTTATATT TTCTGGACTT 4440

AAGTCATAGC CATTCAACTT CATGCTTTTG ACCGAATCAT CTATATAGGT AAGAGATAAA 4500

TAAGAGATAG CTCCTGGACT TTTTGATACG ATTGATTTTA CCGCTCCATT TGAATCCTGC 4560

TCCTGACTTT GCATGGCAGA CTGACCTTCC ATAATGACAG TATCAAAGGT AGCACGAGAG 4620

CCAGAGCCGG CTGCCCGATT GATAACAGAG ATGGGTAAGT CCTTACCACC AACCTCTTTC 4680

CAATTGGTTA CCTCACCTAT GAAGATTTGA CGAAGTTGCT CTGTCGTTAG GTTATCAACA 4740

TCAACCTCCT TATTGACAAT CAGAGCCAAG CCAGCTACCG CGACCTTGTG GTCAACAAGA 4800

GCAGAAGCAT CAATTCCGTC TTTTTCCTCA GCAAATACAT CTGAGTTTCC TATATCAACT 4860

GCCCCAGACT GAACCTGGGA CAAGCCTGTA CCAGAACCTC CCCCTTGGAC ATTGACCGTT 4920

TTTCCAACAT GGATCGTGCC AAATTCATCT GCCGCTACTT CAACCAAGGG TTGCAAGGCA 4980

GTTGAGCCAA CAGCCGTTAT GGATTCTCCA CGATCAATCC AGCTAGCACA GCCTACTAAA 5040

CAAGCCGTCA GCCAAAAAGC GATAAGAGAC AGAGCAAGCT TTTTTCTTTT TTTCACTGTT 5100

TTTCTCCTCG AAAATAATTA TGAATACTGT GAATTTTTTA AGTAGTTCTT TATGAGTTGA 5160

CGCATGAATT CTTACCAAAT TTCTGCGCAA TTGATTATTT ATATAATATA GGCTATATTA 5220

CTCTTTCCTA ACCTCCTTTT TTCATATGTG GATAAAATCT CTTGTCTATC CCTTCCCCCA 5280

TTGTCACCCA TTATAGTCAT TTCGTGTCTC TTTTTCCCCT TTTTAATGCA AGGGAAATTA 5340

CTCTCCTTAG ATGATAATCC AAAAGCTAGA AAGGTATCTC AAACCTCTCT ACTCTCCCAG 5400

ACTAGTTTAC AACTAAAAGG AAAAGATTCT ATTTTATGAG AAATCTAGTT TACAAGCGGT 5460

AAGAACGCTA ATAACTAAAC TTCTTGTACT CTTTGAAAAT CTCTTCAAAC CAGTGTTTTG 5520

AGCTATCTAT GGCTAGCTTC CTAGTTTGCT CTTTGATTTT CATTGAGTAG TAAAACTACA 5580

TGTAATGGCA ATCAAGATAT CAAGAATCAT CCTACTAAAA AAATCCATAC TTTCACTATA 5640

ACATAGAATA AGATATTTGA CTAGCATTTT CATTTGAATC TGAGGCCTTT TGGAAAATAA 5700

TTTTTCAAAA CATTTCCAGT AACCTTTGCA AAGCCCAAGC CATTGCCTTT AACCAAAACT 5760

TGGTACCAAC CATTTGGCAG ACTTTCTGCC AGCTGAACGG TTTCTCCAGC CGCATACTTG 5820

ACAAACGCTT CTTGGCCAAT TTCAACCGAC TGTTCGACCT GACTCGGTTT CAAGGCTAAA 5880

CCAAGAGCGA AACTGGGCTC AAAGCGTTTC TTCTTAAAAG TACCCAGATG CAGTCCATTG 5940

CGAGCAATCT TGAGCTTCCA TAAATCTGGC AAAAGTTCTG GCAAGAGATA AAGCTGGTCT 6000

CCAAAAATCT GCAAGATACC CGGTAGATTG ACCTTCAAAT GGTTTTGGGC AAATTCCTGC 6060

CACAAGGCAA CTTGTTCACG GCTGAGGTTA CTCTTACTTG CCTTAAATTT AGGAGCTGGA 6120

TTGTTACCCT TAAACTGTAG ATGGGCAACA AACTGACCCT CTCCCTTAAA CTGATGAGGA 6180

TACATCCGAG CCGTTTCTGG CAGGTCAATA CCAGCTACCA TTCCATTGAT ATGCTCTACT 6240

GGCAACAAGT CAAAATCATA CTCTTCCAGC AACCAATTGA CAATCTCTTC GTTTTCCTCG 6300

GGTGCCCAGG TACAGGTCGA ATAAACCAGA TGACCACCTT CAGCTAACAT GGTCACTGCA 6360

TCCTCCAGAA TTTCTCTTTG CAAGCTAGCA CATTGACTCG GATAATCTAA GCTCCAATAG 6420

TCCATAGCAT CAGGTTGCTT ACGAAACATT CCTTCACCAG AGCAAGGGGC ATCAAGAACG 6480

ATTAAGTCAA AATAGCCTTT AAAGACCTTG ACCAAGCGGT CGGCAGATTC ATTGGTCACC 6540

ACGACATTTG TCGCTCCAAA ACGCTCCATG TTTTCAACCA AAATCTTAGC CCGTTTGCTT 6600

GAAATTTCAT TGGAAnCAAG TAGCCCCTCC CCTGCTAGAT AGGCTGCCAG TTGAGTTGAT 6660

TTGCCCCCCG GTGCAGCAGC CAAGTCCAAG ACCTTCATAC CAGGACTGGG TTGGGCTACT 6720

TGAGCCACCA TTTGAGCAGC AGGTTCTTGC GAATAAACTA AACCTGTAGC ATGCTCAGGC 6780

GATTTCCCTG AAACCTTCCC ATAGTGGCCC CAAGGGGTTT GAGTAATGGC ATCAGAAAAG 6840

GAAAGTTGCT CTTCTTTTAA GGGATTGACC CGAAAGGCCG AAACCGCTTC CTCCTCAAAA 6900

GAGGCAAGAA AATCTCTTGC CTCATCTCCT AGTATCTCTT TATATTTTTC AACAAATCCT 6960

TCTGGAAATT GCATTTAAGT TCTTTTCCTT TCGTAAATAT AGGACTGAAT TTCCTCCTGC 7020

ATCTCAAGAG GCACCATCAT GACCGGCTGT CTGGTTTGAA AATCAGGAGC TTCACCAAAA 7080

AGGGTCACAA CCCGATAGCC CAGACTTTCC CCTAAAATAC TAGCTGCGGC ATAATCCCAT 7140

GGTTGCAGAT AAGTGAGATA GGTCAACAAA CGCCCTGACA AAATCTTGGC AAAACTAATG 7200

GCCGCACTTC CATAGACACG AACACCAAGA ACCGCTCGGC TCAAATCAGC CAGCCCCCAT 7260

TCATTGGTTT CCAGCATACC ACTATTCCCT GCAATGAGAA AATCTCCAAG TGGTTTAGTT 7320

TTAAAAGGAG CTAGGGACCT ATCATTTAGA CAAACTGGAA ATTCCCCACC ACCGTGGTAA 7380

CAATCCCCTT TGACCACATC ATAAATCAGA CCAAACTGTC CCTGACCATT TTCAAAATAA 7440

GCCATCATAA CAGCAAAATC TTCCTGCTGG GCTACAAAAT TATTGGTACC ATCAATGGGA 7500

TCAATGACCC AAACCTTGCC CTCTTGAACC GAGGCTCGCA GACAACCTTC TTCAGCACAA 7560

ATCTTATCCT CAGGATAACG GGACAAAATC TCACCAACCA AGAGTTCCTG AACTTCTTTG 7620

TCCAGTCTGG TCACCAAATC TGTTGGAGAG GACTTGGTTT CAACACGCAA GTCTTCCTGC 7680

ATATGGTCAA GAATGTACTG ACCTGCTTTC TTAACAAGCT CTTTAGCAAA TTCAAATTTA 7740

CTTTCCAAGA GAAATCTTTC CTTCCCCTTT TTCTTTGGGG 7780

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4820 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double —
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

GTAATGATAT AGGAACACCA GGTGACCTGA TGGGACGTCG TAAGCCTATG AACTACTAGC 60

TGCTAAAGGC TTTAAAGATG GTATGGTACC ATATATCTCA AACCAATACG AAGAAGAAGC 120

CAAACAAAAG GGCAAGACAA TCAATCTCTA CGGTAAAACA AGAGGTTTGG TTACAGATGA 180

CTTGGTTTTG GAAAAGGTAT TTAATAACCA ATATCATACT TGGAGTGAGT TTAAGAAAGC 240

TATGTATCAA GAACGACAAG ATCAGTTTGA TAGATTGAAC AAAGTTACTT TTAATGATAC 300

AACACAGCCT TGGCAAACAT TTGCCAAGAA AACTACAAGC AGTGTAGATG AATTACAGAA 360

ATTAATGGAC GTTGCTGTTC GTAAGGATGC AGAACACAAT TACTACCATT GGAATAACTA 420

CAATCCAGAC ATAGATAGTG AAGTCCACAA GCTCAAGAGA GCAATCTTTA AAGCCTATCT 480

TGACCAAACA AATGATTTTA GAAGTTCAAT TTTTGAGAAT AAAAAATAGT GTCTACTATT 540

AGGAAATAAA GTTTAAAAAG GTGATGAAGA ACAAACCAAG ATTCAAGCAG GAATTCCTAC 600

TGATAATGAA GTAAGTTATG ATCTTATTTA TCAGCAGGAA ACTCTTCCTG CAACAGGTTC 660

ATCAACTTCT GAGCTTACAG CTTTAGGCCT ATTAGCTGTT GGTAGTTTAG TTCTTTTGGT 720

TCATAATATG ACGGGAACAG TTTTTTGCTC CCTCTGAAAA GTCATCATTT GATGGCTTTT 780

TTCTATATAG GGTAAAAGAT AGGGTAAAAG GCTATCATCG GACAAAATAA AGAAGGCATG 840

ATATAATATA AAGTAGATTT CTATGTCATA AAACAAGAAC TGTTTGGACA TCATTCATTT 900

GAAAACTCTC TATGTTCAAA CAATAGTAAA ATAAAATAGG GGATCTAAAT CCTTGCTATG 960

AAAGGAAAAA ACTCAATGGC TACTATTCAA TGGTTTCCTG GTCACATGTC TAAAGCTCGT 1020

CGACAGGTGC AGGAGAATTT AAAATTTGTT GATTTTGTGA CGATTTTAGT AGATGCACGC 1080

TTGCCTCTAT CTAGTCAAAA TCCTATGTTG ACCAAGATTG TTGGTGATAA ACCAAAACTC 1140

TTGATTTTAA ACAAGGCCGA CTTGGCTGAT CCAGCAATGA CCAAGGAATG GCGTCAGTAT 1200

TTTGAATCAC AAGGAATCCA GACGCTAGCT ATCAACTCCA AAGAGCAAGT GACTGTAAAA 1260

GTTGTAACAG ATGCGGCCAA GAAGCTCATG GCTGATAAGA TTGCTCGCCA GAAAGAACGT 1320

GGGATTCAGA TTGAAACCTT GCGTACTATG ATTATCGGGA TTCCAAACGC TGGTAAATCA 1380

ACTCTGATGA ACCGTTTGGC TGGTAAAAAG ATTGCTGTTG TTGGAAACAA GCCAGGGGTC 1440

ACAAAAGGTC AACAATGGCT TAAAACCAAT AAAGACCTGG AAATCTTGGA TACACCGGGG 1500

ATTCTCTGGC CTAAGTTTGA GGATGAAACT GTTGCACTTA AGTTGGCATT GACTGGAGCT 1560

ATCAAAGACC AGTTGCTTCC TATGGATGAG GTTACCATTT TTGGTATCAA TTATTTCAAA 1620

GAACATTATC CAGAAAAGCT GGCTGAACGC TTCAAACAAA TGAAAATTGA AGAAGAAGCG 1680

CCTGTGATTA TTATGGATAT GACCCGCGCC CTCGGTTTCC GTGATGACTA TGACCGTTTT 1740

TACAGTCTCT TCGTGAAGGA AGTCCGTGAT GGCAAACTCG GTAACTATAC CTTAGATACA 1800

TTGGAAGACC TCGATGGCAA CGATTAAAGA AATCAAAGAA TTCCTTGTGA CAGTCAAGGA 1860

GTTAGAAAGC CCTATTTTTT TAGAGCTTGA AAAGGATAAT CGCTCAGGAG TTCAAAAGGA 1920

AATCAGCAAG CGTAAAAGAG CCATTCAAGC TGAATTAGAT GAAAATTTGC GCTTGGAATC 1980

CATGCTTTCT TATGAAAAAG AACTTTATAA GCAAGGATTG ACCTTAATTG CAGGTATTGA 2040

TGAGGTTGGT CGTGGTCCTC TTGCTGGTCC TGTAGTCGCT GCGGCCGTTA TTTTATCTAA 2100

AAATTGTAAG ATTAAAGGTC TCAACGACAG CAAGAAAATT CCTAAAAAGA AACATCTGGA 2160

GATTTTCCAA GCCGTTCAAG ACCAAGCCTT GTCGATTGGA ATTGGTATCA TAGATAATCA 2220

GGTCATCGAC CAAGTCAACA TCTATGAAGC AACCAAACTA GCCATGCAAG AAGCAATCTC 2280

CCAGCTCAGC CCTCAACCAG AGCACCTTTT GATTGATGCC ATGAAACTGG ACTTGCCCAT 2340

TTCACAAACC TCCATTATCA AAGGAGATGC CAACTCCCTC TCTATCGCAG CAGCATCTAT 2400

AGTAGCCAAG GTAACACGTG ATGAATTGCT GAAAGAATAC GATCAGCAGT TCCCTGGCTA 2460

TGATTTCGCT ACTAATGCAG GATATGGCAC AGCTAAACAT CTGGAAGGCC TCACAAAACT 2520

AGGAGTTACC CCAATTCACC GAACCAGCTT TGAACCCGTT AAATCACTGG TTTTAGGTAA 2580

AAAAGAAAGT TAATTGAAAG GAAATAACAT GGAGGAACAG TCGGAAATAG TCCGTTCTAA 2640

GAAAGAATTC GCCTTTGCAT CCAGCACTAT ACTATCCCAA GTTGGTCGAG GAATCATTGT 2700

CGGCCTCATC GTTGGAATTA TCGTCGGATC CTTTCGTTTC TTAATTGAAA AGGGCTTCCA 2760

CCTGATACAA GGAGTTTATC AAGATCAAGG GTACTTAGTG CGCAATCTTT TTGTACTGGT 2820

TTTGTTTTAT ATACTCATCT GTTGGCTCAG TGCCAAACTA ACACGGTCAG AAAAAGATAT 2880

TAAAGGCTCA GGAATTCCTC AAGTCGAAGC CGAACTGAAA GGCCTCATGT CCCTCAACTG 2940

GTGGGGCATT CTTTGGAAAA AATATGTGCT AGGTATTCTT GCTATTGCCA GTGGACTCAT 3000

GCTGGGTCGA GAGGGACCCA GCATTCAACT TGGAGCAGTT GGTGGTAAAG GAATTGCCAA 3060

GTGGCTCAAA TCCAGTCCAG TAGAGGAACG TTCCTTGATT GCCAGTGGAG CTGCAGCAGG 3120

TTTAGCCGCA GCCTTTAATG CTCCTATTGC AGCACTTCTC TTTGTTGTAG AAGAAGTCTA 3180

TCACCATTTT TCGCGCTTTT TCTGGGTCTC AACTCTAGCA GCCAGCATCG TAGCAAACTT 3240

TGTGTCTCTA CTCATGTTCG GTTTGACACC AGTATTGGAT ATGCCAGATA ACATTCCTCC 3300

CATGACCCTA GATCAGTATT GGATATATCT CGTCATGGGA ATTTTCCTTG GATTTTCAGG 3360

TTTTCTCTAT GAGAAAGCTG TATTAAACGT TGGAAGAGTT TATGACTTGA TTGGTCAAAA 3420

AATCCATTTG GATAGGGCTT ATTATCCCAT CTTGGCTTTT ATCCTTATCA TACCAGTCGG 3480

AATCTTCTTA CCTCAAATCA TTGGTGGCGG AAATCAGCTT GTCCTTTCTT TAACTGAACA 3540

AAATTTTAGT TTCCAAGTTT TATTAGCTTA CTTTTTAATC CGCTTTATTT GGAGTATGAT 3600

TAGCTATGGA AGTGGACTGC CAGGAGGAAT TTTCCTCCCC ATTTTAGCTC TTGGTTCTTT 3660

GCTTGGTGCC TTAGTTGGTG TTATCTGTGT CAATCTTGGA CTTGTCAGTC AAGAGCAATT 3720

CCCTATATTT GTCATTCTAG GAATGAGTGG CTATTTTGGA GCCATATCAA AAGCTCCCTT 3780

AACCGCTATG ATCCTCGTAA CTGAGATGGT AGGAGATATT CGCAACCTTA TGCCACTTGG 3840

TCTTGTCACT CTTGTTTCTT ATATTATCAT GGATTTGCTC AAAGGTACGC CAGTCTATGA 3900

AGCCATGCTG GAAAAAATGC TTCCAGAAGA AGTATCTAGC GAAGGAGAAG TTACACTTAT 3960

CGAAATACCA GTTTCTGATA AAATTGCTGG GAAACAAGTT CATGAACTCA ACTTACCACA 4020

CAACGTCCTC ATCACAACTC AAGTCCATAA TGGCAAGAGC CAAACAGTTA ACGGCTCAAC 4080

CAGAATGTAT CTGGGTGATA TGATTCACCT GGTTATTCCA AAAAGTGAAA TTGGAAAAGT 4140

CAAAGATTTG TTGTTGTAGT ATGAGTATTT ACATAATTTA TGTTATGTAA ATGATCAGTT 4200

TGATTTATTT AGAAAACCGA TTCTCAGGAA TGAGATCGGT TATTTTTTAC TGATGAGGAA 4260

TTTTACATAT AAATAATTGA ACTTTATTAA AAATAAGACT ATAATTAAGT TAGAAATGAT 4320

AAAGTATAAA GCTAGAAAGG AGTTTACTGT ATCAAATCTG TACAGTAAGA TTAAAATCAT 4380

GAAAAAGAAA ACAATAGCAA TTATATAGAG AAATGAAATA GAAATAGGAT AAAACAATCA 4440

GGACAATCAA ATCAATTTCT AGCAATGTTT TAGAAGTCCA GATGTACTAT TCTAGTTTCA 4500

ATCTATTATA CAATGTGTTT TGTATCTCAT AGCTCCTTAT ATAGCTCTTC AGTTATGTAG 4560

TATTAACAGA AGTTTAGTGG GTGAGATTTT TATTATTTTC CTTATTCTGT TTTGTTTGTA 4620

GGTCTAAGTC TTTTTATCAC TTTGAAAAAC TCCTATAACA TCTTTCCGAA AAACTATAAT 4680

TTTCTTGAAA AATATACAAG TCTATGCTAT ACTACTAGTA TACTTACTTA TGGAGAAAAT 4740

ACATGAAACG TGAGATTTTA CTGGAACGAA TCGACAAACT AAAACAACTC ATGCCCTGGT 4800

AAGTTCTGGA ATACTACCAA 4820

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21338 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

CTACGACATC ATGATTAACA GTCATGCGCT ACTACCAACT GAGCTATGGC GGATAAAATA 60

GTCCGTACGG GATTCGAACC CGTGTTACCG CCGTGAAAAG GCGGTGTCTT AACCCCTTGA 120

CCAACGGACC TTCTATCTGT AGCAGATATA ACCATTATAT CAATTTCTTG CTAATTGTCA 180

ATCACTTTTG AGATTTTTTC TCTAAAATAT CTTTTAATTT TCTAATTTTT AATCTTGAAA 240

TAGGACAACG ATGGTCTTCA TAGAAAACAA TTTCTAAGTT TTTTCGATCA ATTTCTCTGA 300

TATTACCTAT ATTTACCAAA AATGACTTGT GAGGAGAATA AAATCGCTGA GTATGTTTGT 360

CCTTTTCCTG AATATCTGTC ATGGTACCAT AAAACTCTTT TGCAAAATTC TTACCAATAA 420

TGCGCAATTT ATGAGATACC CCTGTTGTTT CAATATACAA AATATCATGG TAAGGAATTT 480

TTAAATCATT TCCCTTGTAA TTGTAGTCGA AATAATCTAC AACATCTTCA TTTTCAAGTA 540

ACATACTCTT CGTGTAGAAG ATATTTTGCT CAATTCTCTT CTTAAACATC TCATCATTGA 600

TATCCTTATC AACAAAATCT AGGGCTGATA CCTGGTATTT ATAGGTTAGA GTCGCAAACT 660

CTGATCGACT AGTGATAAAG ACGATAATAG CGTAAGGATT GTAATGACGA ATGAGCTGAG 720

CCACTTCAAA TCCCTTTTTC TCAATTCCAT GAATATCGAT ATCTAGGAAA TAAAGCTGAT 780

TTACTTCATC ATTTTCAATG TATTCTTCAA ATTCACGGAC TTTTCCCGTT GTCTTGTATG 840

ATATTGGAAT ATTCGATTCT TTCGAAATTT CATCCAATAT TCTCTCTAGT CTCACTTGAT 900

GTTCAATAAC ATCTTCTAAA ATTAAAACTT TCATTCAAAT TCCCTCTTAA ATCTAATGAT 960

TTGTCTAAAT GTACTGCCTT CCATCTCTGT TTCTAAAATA ATATTGTTGT ACTTATCTAG 1020

TAGTTCTTTC ACATTATTTA ATCCGACTCC GCGATTTCTT CCCTTAGTGG AGAATCCTAA 1080

GGCAAATAGA TCTCCTGAAG GAGTCATCGT CATTTTACAT GAATTCTGAA TCACAATAAC 1140

TGTTTCAGTT TCCATCTTAA TAACTGCTAC TTCCATCTGC TTTTTATAGC TATCAGCCGA 1200

TCCTTCGACA GCATTATTCA ATAAAACGCT CATGATACGA ACCAAATCCA ATAGTTCAAT 1260

TGGAAGCTTG GTAATCGTAT CTTTTACTTC CAGTGTAAAC TCTACACCAT TATTTCGAGC 1320

ATAGACAATT GACTGAGCAA CCAAACTTCG TAAAGCTGAG TCTTCTATGT TGTTCAAATC 1380

AAAGTAAGTG TACTTATCTG AACGCAATTT ATGATTTGCT TTGACTAAAA CTTCATTGTA 1440

AATTCTGTCA ATTTCCTGTA AATTACCACT GTCAATTGCC ATCTGCATGC TGACAAGCAT 1500

TCCAGCATAA TCATGTCGAA AACCACGGAT TTCATTATAC AGACCAACAA TTTCATCTGT 1560

GTAATTCTGT AAATGTTTCT GTTCAAATTT CTTCTGCTTC AAAGCAATCT CTTTCTCCAT 1620

TTGAACTTTA TGAGAATTCA TTGCAAAGAA GGTCAAAAGG AGAGAGATAA AGACAATAGA 1680

TGACAAAATA CTTCCAAAAC TATTCAAATG TTTAATCGTA CTTACCATAT CTGAAACGAA 1740

AGATACAATA TGTAGCAATA GTAAAGCAAA AAATACTTTT TTCAAGAAAG GATAAAGGTA 1800

GTCCTTGTCA AAATAGGCTA GTTCCAAATG GAAATAGTAA ATGATTTTTA ATGTAACAAA 1860

ATAGGTTAAC ACCGTCACAA CGAAAAAGAA TGGGAAATGA TATTGTAAAA CAAAATTATC 1920

TCCTGTTATA GAGGAGAAAA TTACGGACAG AAAGTTATGA GTGCTCTCAT ATAAAAGAGA 1980

TAGTAGTAAA CTTAGGAATA GTCCTCTATC CCTCTCATAC TGTTTCATCC ATCGAAAATA 2040

GGAATATAAG CCCAAAGGAA ATAAAAATCT TTCAATCCCT ATTTTATCTA AATATAGAAG 2100

ATAAAAGGAA AATTCAAGTA CTATTTCAGT TAGTAATGTA TAAGCACCAA AAACGTATAA 2160

TTCTTTTCTA TTTATTCGAC CTTTACAAAT TAAACGGTAA CTGTGACTAA TAATTAAAAA 2220

ATGAACAATA ACTGTCCCAA ATCCAAGTAA ATCCATTACT CTTTCTCCTT ATTTCATTAC 2280

TTTTTTCGTA GGAAAAGAAA ATCAAGGATG ATTCTTGAAA TCCTCATCTC CCCACCTTTA 2340

ATCTTTTGTA AGTCTTTTTC CTTCAAAGCT ACAAACTGTT CCAATTTAAC TGTGTTTTTC 2400

ATAATAAAAT CTCCTAAAAT GTTTTTTCTT GTAAGCTAAC TTACAAAAAC CATTATACAA 2460

AATGGAATTT CGTTTTAGAT AAAATTCTCT CAACTGTCAT TTTTTTCTCC CAAAGTGTAC 2520

TTTTTTAAGA AAAAAGCCGG GAAAATTCCC AGCTTTGCTA TTATATTGAT CCCAGCAGGA 2580

TTCGAACCTG CGACCGTTCG CTTAGAAGGC GAATGCTCTA TCCAGCTGAG CTATGAGACC 2640

TAATACAATT ATTCTACCAA AAATTCAATT AAAAGTCAAT TTTCTATTTA TGGTAGGGGA 2700

ATCCCTGCTG AATCGTAAAA GCGCGATAGA TTTGTTCAAC AAGAACTAGT CTCATTAACT 2760

GATGGGGTAA GGTTAGGCGA CCAAAACTGA CAGAAAGATT GGCTCTATTT TTTACAGATG 2820

ATGATAATCC TAAACTTCCC CCAATAATAA AAGTAAGAGT AGAAAATCCT TTTATAGAAG 2880

TTTCTTCTAA CTGCTTACTA AATTCTTCTG AGAAGAAAGT TTTCCCTTCA ATGGCTAACA 2940

CAATAACGAA ATCACGGTCA GCAATTTTTG ATAAAATTCT CTGACCTTCT ATTTCTAAAA 3000

TCTTTTGATT TTCTGATTCA CTGGCCTTAT CTGGTGTTTT TTCATCTGAT AACTCAATCA 3060

TTTCAAACTT AGCAAATCTA GAAATTCGTT TTGAATACTC TGCGATACCA TCTTTTAAAT 3120

ACTTTTCTTT CAGTTTCCCA ACTGTTACAA CTTTAATTTT CATGACTCTA TTCTAACATA 3180

TTCTCTATTT TTTCACATCT TATTCACAAA ATAAAAAAATA GATTTCAATT AAGAAAATCA 3240

CAATTTCAAA AGAGTTATCC ACAGTTTGTG TAAAACTTTT GTGTTTAAGT TATAATTAAG 3300

CTAGTCAGTT TATACTTTCA GTAATTCAAA CATATGGAGG CAAATATGAA ACATCTAAAA 3360

ACATTTTACA AAAAATGGTT TCAATTATTA GTCGTTATCG TCATTAGCTT TTTTAGTGGA 3420

GCCTTGGGTA GTTTTTCAAT AACTCAACTA ACTCAAAAAA GTAGTGTAAA CAACTCTAAC 3480

AACAATAGTA CTATTACACA AACTGCCTAT AAGAACGAAA ATTCAACAAC ACAGGCTGTT 3540

AACAAAGTAA AAGATGCTGT TGTTTCTGTT ATTACTTATT CGGCAAACAG ACAAAATAGC 3600

GTATTTGGCA ATGATGATAC TGACACAGAT TCTCAGCGAA TCTCTAGTGA AGGATCTGGA 3660

GTTATTTATA AAAAGAATGA TAAAGAAGCT TACATCGTCA CCAACAATCA CGTTATTAAT 3720

GGCGCCAgCA AAGTAGATAT TCGATTGTCA GATGGGACTA AAGTACCTGG AGAAATTGTC 3780

GGAGCTGACA CTTTCTCTGA TATTGCTGTC GTCAAAATCT CTTCAGAAAA AGTGACAACA 3840

GTAGCTGAGT TTGGTGATTC TAGTAAGTTA ACTGTAGGAG AAACTGCTAT TGCCATCGGT 3900

AGCCCGTTAG GTTCTGAATA TGCAAATACT GTCACTCAAG GTATCGTATC CAGTCTCAAT 3960

AGAAATGTAT CCTTAAAATC GGAAGATGGA CAAGCTATTT CTACAAAAGC CATCCAAACT 4020

GATACTGCTA TTAACCCAGG TAACTCTGGC GGCCCACTGA TCAATATTCA AGGGCAGGTT 4080

ATCGGAATTA CCTCAAGTAA AATTGCTACA AATGGAGGAA CATCTGTAGA AGGTCTTGGT 4140

TTCGCAATTC CTGCAAATGA TGCTATCAAT ATTATTGAAC AGTTAGAAAA AAACGGAAAA 4200

GTGACGCGTC CAGCTTTGGG AATCCAGATG GTTAATTTAT CTAATGTGAG TACAAGCGAC 4260

ATCAGAAGAC TCAATATTCC AAGTAATGTT ACATCTGGTG TAATTGTTCG TTCGGTACAA 4320

AGTAATATGC CTGCCAATGG TCACCTTGAA AAATACGATG TAATTACAAA AGTAGATGAC 4380

AAAGAGATTG CTTCATCAAC AGACTTACAA AGTGCTCTTT ACAACCATTC TATCGGAGAC 4440

ACCATTAAGA TAACCTACTA TCGTAACGGG AAAGAAGAAA CTACCTCTAT CAAACTTAAC 4500

AAGAGTTCAG GTGATTTAGA ATCTTAATTG ACATCTATGT AAAGAAAGCT TTACATAAGA 4560

GAAAAGATGT GTTAGTGTAG AATCATGGAA AAATTTGAAA TGATTTCTAT CACAGATATA 4620

CAAAAAAATC CCTATCAACC CCGAAAAGAA TTTGATAGAG AAAAACTAGA TGAACTAGCA 4680

CAGTCTATCA AAGAAAATGG GGTCATTCAA CCGATTATTG TTCGTCAATC TCCTGTTATT 4740

GGTTATGAAA TCcTTGCAGG AGAGAGACGC TATCGGGCTT CACTTTTAGC TGGTCTACGG 4800

TCTATCCCAG CTGTTGTTAA ACAGATTTCA GACCAAGAGA TGATGGTCCA GTCCATTATT 4860

GAAAATTTAC AGAGAGAAAA TTTAAACCCA ATAGAAGAAG CACGCGCCTA TGAATCTCTC 4920

GTAGAGAAAG GATTCACCCA TGCTGAAATT GCAGATAAGA TGGGCAAGTC TCGTCCATAT 4980

ATCAGCAACT CCATTCGTTT ACTTTCCTTG CCAGAACAGA TTCTTTCAGA AGTAGAAAAT 5040

GGCAAACTAT CACAAGCCCA TGCGCGTTCC CTAGTTGGGT TAAATAAGGA ACAACAAGAC 5100

TATTTCTTTC AACGGATTAT AGAAGAAGAT ATTTCTGTAA GGAAATTAGA AGCTCTTCTG 5160

ACAGAGAAAA AACAAAAGAA ACAGCAAAAA ACTAATCATT TCATACAAAA TGAAGAAAAA 5220

CAGTTAAGAA AACTACTCGG ATTAGATGTA GAAATTAAAC TATCTAAAAA AGACAGTGGA 5280

AAAATCATTA TTTCTTTTTC AAATCAAGAA GAATATAGTA GAATTATCAA CAGCCTGAAA 5340

TAAGGCTGTT CTTTTATTTT TTTATCTCAC AAGGTTATCC ACTATGTTTT TCGATAAAAA 5400

GCTTAATAAA TCAATAATTT CTTCTTTTAT CCCCAACCTG TGGATAAAGT TTGGTAACAT 5460

TGTGGATTAT TTTTCACAGC TTGTGGAAAA TTCTTGCTAT CTATGGTAAA ATATCTCTAG 5520

TATTAAACTT TTAAATAGTA AAGGAGGAGA AAGGATTGAA AGAAAAACAA TTTTGGAATC 5580

GTATATTAGA ATTTGCACAA GAAAGACTGA CTCGATCCAT GTATGATTTC TATGCTATTC 5640

AAGCTGAACT CATCAAGGTA GAGGAAAATG TTGCCACTAT ATTTCTACCT CGCTCTGAAA 5700

TGGAAATGGT CTGGGAAAAA CAACTAAAAG ATATTATTGT AGTAGCTGGT TTTGAAATTT 5760

ATGACGCTGA AATAACTCCC CACTATATTT TCACCAAACC TCAAGATACG ACTAGCTCAC 5820

AAGTTGAAGA AGCTACAAAT TTAACTCTTT ATAACTATAG TCCAAAGTTA GTATCTATTC 5880

CTTATTCAGA TACGGGATTA AAAGAAAAGT ATACCTTTGA TAACTTTATT CAAGGGGATG 5940

GAAATGTTTG GGCTGTATCA GCCGCTTTAG CTGTCTCTGA AGATTTGGCT CTGACCTATA 6000

ACCCTCTTTT TATCTATGGA GGACCAGGCC TTGGTAAGAC TCACTTATTA AACGCTATTG 6060

GAAATGAAAT TCTAAAAAAT ATTCCTAATG CGCGTGTTAA ATATATCCCT GCCGAAAGCT 6120

TTATTAATGA CTTTCTTGAT CACCTAAGAC TTGGGGAAAT GGAAAAGTTT AAAAAGACCT 6180

ATCGTAGTCT TGATCTTTTG TTAATCGATG ATATCCAGTC ACTCAGCGGA AAAAAAGTCG 6240

CAACTCAGGA AGAATTTTTC AATACCTTTA ACGCCCTTCA TGACAAGCAA AAACAGATTG 6300

TCCTAACGAG TGATCGTAGT CCAAAACATC TAGAAGGGCT CGAGGAGAGG CTTGTCACGC 6360

GTTTTAGTTG GGGATTGACA CAAACTATCA CCCCCCCTGA CTTTGAAACA CGTATTGCCA 6420

TTTTACAAAG TAAGACGGAA CATTTAGGCT ACAATTTCCA AAGTGATACT CTAGAATACC 6480

TAGCTGGGCA ATTTGATTCA AATGTTCGAG ATCTTGAGGG AGCCATCAAC GACATCACTT 6540

TAATTGCCAG AGTAAAAAAA ATCAAGGATA TCACTATTGA TATTGCTGCA GAAGCCATTA 6600

GAGCCCGCAA ACAAGATGTT AGCCAAATGC TCGTCATCCC AATTGATAAA ATCCAAACTG 6660

AAGTTGGTAA CTTTTATGGT GTTAGTATCA AAGAAATGAA GGGAAGTAGA CGCCTTCAAA 6720

ATATTGTTTT GGCCCGTCAA GTAGCCATGT ATTTATCTAG AGAACTAACA GATAATAGTC 6780

TTCCAAAAAT TGGGAAGGAA TTTGGGGGAA AAGATCATAC CACAGTCATT CATGCCCATG 6840

CCAAAATAAA ATCTTTGATT GATCAAGACG ATAATTTACG TTTAGAAATT GAATCAATCA 6900

AAAAGAAAAT CAAATAATTT GTGGATAACT TTTAGTTTTT TATCTTTTTT ATCCACATTT 6960

TTTAAACAAG CTAAAAAACT TGATATGACT TGTTTAAAGG CTGTTTTCCA CAGATTTCAC 7020

AGACTCTATT ATTACTATTA TCTTTCTAAT ACTAAAAATA AATAAAGGAG AATCCATGAT 7080

TCATTTTTCA ATTAATAAAA ATTTATTTCT ACAAGCATTA AATACTACTA AGAGAGCTAT 7140

TAGTTCTAAA AATGCCATTC CTATTTTATC AACAGTAAAA ATTGACGTGA CCAATGAAGG 7200

TATTACTTTA ATTGGTTCAA ATGGTCAAAT TTCAATTGAA AATTTTATTT CTCAAAAAAA 7260

TGAAGATGCT GGTTTGTTAA TTACTTCTTT AGGTTCGATC CTTCTTGAAG CTTCTTTCTT 7320

TATCAATGTA GTATCTAGTT TACCTGATGT AACTCTTGAT TTTAAAGAAA TTGAACAAAA 7380

TCAAATTGTT TTAACCAGTG GCAAATCAGA AATTACCCTA AAAGGAAAAG ATAGCGAACA 7440

ATATCCACGA ATCCAAGAAA TTTCAGCAAG CACTCCTTTA ATACTTGAAA CAAAATTACT 7500

CAAGAAAATT ATTAATGAAA CAGCCTTTGC TGCAAGTACA CAAGAGAGTC GTCCGATTTT 7560

AACAGGTGTC CACTTCGTAT TGAGTCAACA CAAAGAGTTA AAAACAGTTG CAACAGACTC 7620

TCATCGCCTA AGCCAGAAAA AATTGACTCT TGAAAAAAAT AGTGATGATT TTGATGTCGT 7680

AATTCCTAGC CGTTCTCTAC GCGAATTTTC AGCGGTATTT ACAGATGATA TCGAAACTGT 7740

AGAGATTTTC TTTGCCAATA ACCAAATCCT CTTTAGAAGC GAAAATATTA GCTTCTATAC 7800

TCGTCTCCTA GAAGGAAACT ATCCTGATAC AGATCGCTTG ATTCCAACAG ACTTTAACAC 7860

TACTATTACT TTTAATGTGG TAAACTTACG CCAGTCAATG GAGCGTGCCC GTCTTTTATC 7920

AAGTGCGACT CAAAATGGTA CTGTGAAACT TGAAATTAAG GATGGGGTTG TTAGCGCCCA 7980

TGTTCACTCT CCAGAAGTTG GTAAAGTAAA CGAAGAAATC GATACTGATC AGGTTACTGG 8040

TGAAGATTTG ACCATTAGTT TCAACCCAAC TTACTTGATT GATTCTCTTA AAGCTTTAAA 8100

TAGCGAAAAG GTGACTATTA GCTTTATCTC AGCTGTTCGT CCATTTACTC TTGTGCCAGC 8160

AGATACTGAC GAAGACTTCA TGCAGCTCAT TACACCAGTT CGTACAAATT AAGTGAAAGA 8220

GGTTGAGCCT GGCTCGCCTC TTTTATGATA TAATCGAAAA AGAAAAGGAG AGTAGTATGT 8280

ATCAAGTTGG AAATTTTGTT GAGATGAAAA AATCACACGC TTGTACAATC AAGTCGACTG 8340

GTAAAAAGGC TAATCGTTGG GAAATTACAC GTGTAGGAGC AGATATCAAA ATAAAATGTA 8400

GTAATTGTGA GCATGTTGTC ATGATGGGGC GATATGATTT TGAGCGAAAA ATGAATAAAA 8460

TTATTGACTG AGAACCCTTA GTTAGAGGGT TAGCACTTTA TCCCTTTTTG TGTTATAATA 8520

TTAGGGATTG AAATGAAAAC GGAGAATGAG AAATATGGCT TTGACAGCAG GTATCGTTGG 8580

TTTGCCAAAC GTTGGTAAAT CAACACTATT TAATGCAATT ACAAAAGCAG GAGCAGAGGC 8640

AGCAAACTAC CCATTTGCGA CGATTGATCC AAATGTTGGA ATGGTGGAAG TTCCAGATGA 8700

ACGCCTACAA AAACTAACTG AAATGATAAC TCCTAAAAAG ACAGTTCCCA CAACATTTGA 8760

ATTTACAGAT ATTGCAGGGA TTGTAAAAGG AGCTTCAAAA GGAGAGGGGC TAGGGAATAA 8820

ATTCTTGGCC AATATTCGTG AAGTAGATGC GATTGTTCAC GTAGTTCGTG CTTTTGATGA 8880

TGAAAATGTA ATGCGCGAGC AAGGACGTGA AGACGCCTTT GTAGATCCAC TTGCAGATAT 8940

TGATACCATT AATCTGGAAT TGATTCTTGC TGACTTAGAA TCAGTGAACA AACGATATGC 9000

GCGTGTAGAA AAGATGGCAC GTACGCAAAA AGATAAAGAA TCAGTAGCAG AATTCAATGT 9060

TCTTCAAAAG ATTAAACCAG TCCTAGAAGA CGGGAAATCA GCTCGTACCA TTGAATTTAC 9120

AGATGAGGAA CAAAAGGTTG TCAAAGGTCT TTTCCTTTTG ACGACTAAAC CAGTTCTTTA 9180

TGTAGCTAAT GTGGACGAGG ATGTGGTTTC AGAACCTGAC TCTATCGACT ATGTCAAACA 9240

AATTCGTGAA TTTGCAGCGA CAGAAAATGC TGAAGTAGTC GTTATTTCTG CGCGTGCTGA 9300

GGAAGAAATT TCTGAATTGA ATGATGAAGA TAAAAAAGAG TTTCTTGAAG CCATTGGTTT 9360

GACAGAATCA GGTGTAGATA AGTTGACGCG TGCAGCTTAC CACTTGCTTG GATTGGGAAC 9420

TTACTTCACA GCTGGTGAAA AAGAAGTTCG CGCTTGGACT TTCAAACGTG GTATGAAGGC 9480

TCCTCAAGCA GCTGGTATTA TCCACTCAGA CTTTGAAAAA GGCTTTATTC GTGCAGTAAC 9540

CATGTCATAT GAAGATCTAG TGAAATACGG ATCTGAAAAG GCCGTAAAAG AAGCTGGACG 9600

CTTGCGTGAA GAAGGAAAAG AATATATCGT TCAAGATGGC GATATCATGG AATTCCGCTT 9660

TAATGTCTAA AAATTAATAA ATGGTGTCAA TTAGGTTGGA AAAAAATTCC AACCCTTTTG 9720

GCTTTTGAAA GGAAAAATAA ATGACCAAAT TACTTGTAGG CTTGGGAAAT CCAGGGGATA 9780

AATATTTTGA AACAAAACAC AATGTTGGTT TTATGTTGAT TGATCAACTA GCGAAGAAAC 9840

AGAATGTCAC TTTTACACAC GATAAGATAT TTCAAGCTGA CCTAGCATCC TTTTTCCTAA 9900

ATGGAGAAAA AATTTATCTG GTTAAACCAA CGACCTTTAT GAATGAAAGT GGAAAAGCAG 9960

TTCATGCTTT ATTAACTTAC TATGGTTTGG ATATTGACGA TTTACTTATC ATTTACGATG 10020

ATCTTGACAT GGAAGTTGGG AAAATTCGTT TAAGAGCAAA AGGCTCAGCA GGTGGTCATA 10080

ATGGTATCAA GTCTATTATT CAACATATAG GAACTCAGGT CTTTAACCGT GTTAAGATTG 10140

GAATTGGAAG ACCTAAAAAT GGTATGTCAG TTGTTCATCA TGTTTTGAGT AAGTTTGACA 10200

GGGATGATTA TATCGGTATT TTACAGTCTG TTGACAAAGT TGACGATTCT GTAAACTACT 10260

ATTTACAAGA GAAAAATTTT GAGAAAACAA TGCAGAGGTA TAACGGATAA ATGGTGACCT 10320

TATTAGATTT ATTCTCAGAA AATGATCAGA TTAAAAAATG GCATCAAAAT TTAACAGATA 10380

AGAAAAGACA ACTAATACTT GGTTTATCAA CATCTACTAA GGCTCTTGCA ATTGCAAGCA 10440

GTTTAGAAAA AGAAGATAGG ATTGTGTTAT TGACGTCAAC TTATGGAGAA GCAGAAGGAC 10500

TTGTTAGTGA TCTTATTTCT ATCTTGGGTG AGGAACTCGT CTATCCATTT TTGGTAGATG 10560

ATGCTCCTAT GGTGGAGTTT TTGATGTCTT CACAGGAAAA AATTATTTCA CGGGTTGAAG 10620

CCTTGCGTTT TTTGACTGAT TCATCTAAGA AAGGGATTTT AGTTTGTAAT ATCGCAGCAA 10680

GTCGATTGAT TTTACCGTCT CCCAATGCAT TCAAAGATAG TATTGTAAAA ATCTCAGTTG 10740

GTGAAGAATA TGATCAACAC GCGTTTATCC ATCAGTTAAA GGAAAATGGC TATCGAAAAG 10800

TTACTCAAGT ACAAACTCAG GGCGAATTTA GTCTTCGAGG AGATATTTTA GATATTTTTG 10860

AAATATCCCA GTTAGAACCT TGTCGAATTG AGTTTTTTGG TGATGAAATT GATGGTATCA 10920

GGTCATTTGA AGTAGAAACA CAATTATCGA AAGAAAATAA GACAGAACTC ACTATCTTTC 10980

CAGCTAGTGA TATGCTTTTG AGAGAAAAGG ATTATCAACG AGGACAGTCA GCTTTAGAAA 11040

AACAAATTTC AAAAACTTTA TCACCTATTT TGAAATCATA CCTAGAAGAA ATTCTTTCAA 11100

GTTTTCACCA AAAACAAAGT CATGCAGACT CTCGGAAGTT TTTATCTTTG TGCTATGATA 11160

AGACATGGAC TGTCTTTGAT TATATTGAAA AAGATACTCC AATATTCTTT GATGATTATC 11220

AAAAATTGAT GAATCAGTAT GAAGTCTTTG AAAGAGACTT AGCGCAGTAC TTTACAGAAG 11280

AATTACAGAA TAGTAAAGCA TTTTCTGATA TGCAGTATTT TTCTGATATT GAACAAATCT 11340

ATAAAAAACA AAGTCCAGTG ACCTTTTTCT CTAATCTTCA AAAGGGTTTA GGAAATCTCA 11400

AATTTGACAA AATTTATCAA TTCAATCAAT ATCCTATGCA GGAATTTTTC AATCAGTTTT 11460

CTTTTCTAAA AGAAGAAATT GAACGATATA AAAAAATGGA TTACACCATT ATTCTGCAGT 11520

CTAGCAATTC AATGGGAAGT AAAACATTGG AGGATATGTT AGAGGAATAT CAGATTAAAT 11580

TGGATTCTAG AGATAAGACA AATATCTGTA AAGAATCTGT AAACTTAATA GAGGGTAATC 11640

TCAGACATGG TTTTCATTTT GTAGATGAAA AGATTTTATT GATAACTGAA CATGAGATTT 11700

TTCAAAAGAA ATTAAAGCGT CGTTTTCGAA GACAACATGT TTCAAATGCA GAGAGATTAA 11760

AAGATTACAA TGAACTTGAA AAAGGGGACT ATGTTGTCCA TCATATCCAT GGGATTGGTC 11820

AATATCTAGG AATTGAAACC ATTGAAATCA AGGGAATTCA TCGCGATTAT GTCAGTGTCC 11880

AATACCAAAA TGGTGATCAA ATTTCTATCC CCGTGGAACA GATTCATCTA CTGTCCAAAT 11940

ATATTTCAAG TGATGGTAAA GCTCCAAAAC TCAATAAATT AAATGACGGT CATTTTAAAA 12000

AGGCCAAGCA AAAGGTTAAG AACCAGGTAG AGGATATAGC TGATGATTTA ATCAAACTCT 12060

ACTCTGAACG TAGTCAGTTG AAGGGTTTTG CTTTCTCAGC TGATGATGAT GATCAAGATG 12120

CCTTTGATGA TGCTTTCCCT TATGTTGAAA CGGATGATCA ACTTCGTAGT ATTGAGGAAA 12180

TCAAGAGGGA TATGCAGGCT TCTCAGCCAA TGGATCGACT TTTAGTTGGG GATGTTGGTT 12240

TTGGAAAGAC TGAAGTTGCT ATGCGTGCAG CCTTTAAAGC AGTCAATGAT CACAAACAGG 12300

TTGTCATTCT AGTTCCGACG ACGGTTTTAG CGCAACAGCA CTATACGAAT TTTAAGGAAC 12360

GATTCCAAAA TTTTGCAGTT AATATTGATG TGTTGAGTCG CTTTAGAAGT AAAAAAGAGC 12420

AGACTGCAAC ACTTGAAAAA TTGAAAAACG GTCAAGTCGA TATTTTGATT GGAACACATC 12480

GTGTTTTGTC AAAAGATGTT GTGTTTGCTG ATTTGGGCTT GATGATTATT GATGAGGAAC 12540

AGCGATTTGG TGTCAAGCAT AAGGAAACTT TGAAAGAACT GAAGAAACAA GTGGATGTCC 12600

TAACCTTGAC CGCTACGCCA ATCCCTCGTA CCCTCCATAT GTCTATGCTG GGAATCAGAG 12660

ATTTATCTGT TATTGAAACT CCGCCGACTA ATCGCTATCC TGTTCAGACC TATGTTTTGG 12720

AAAAGAATGA TAGTGTCATT CGTGATGCTG TCTTGCGTGA AATGGAGCGT GGAGGTCAAG 12780

TTTATTATCT TTACAACAAA GTTGACACAA TTGTTCAGAA GGTTTCAGAA TTACAGGAGT 12840

TGATTCCGGA GGCTTCGATT GGATATGTTC ATGGTCGAAT GAGTGAAGTC CAGTTGGAAA 12900

ATACTCTATT AGACTTTATT GAGGGACAAT ACGATATCTT GGTGACGACT ACTATTATTG 12960

AGACAGGGGT GGACATTCCA AATGCTAATA CTTTATTTAT TGAAAATGCG GACCATATGG 13020

GCTTGTCAAC CTTATATCAG TTAAGAGGAA GAGTCGGTCG TAGTAATCGT ATTGCTTATG 13080

CTTATCTCAT GTATCGTCCA GAAAAATCAA TCAGTGAAGT CTCTGAAAAG AGATTAGAAG 13140

CGATTAAAGG ATTTACAGAA TTGGGCTCTG GCTTTAAGAT TGCAATGCGA GATCTTTCGA 13200

TTCGTGGAGC AGGAAATCTT TTAGGAAAAT CCCAGTCTGG TTTCATTGAT TCTGTTGGTT 13260

TTGAATTGTA TTCGCAGTTA TTAGAGGAAG CTATTGCTAA ACGAAACGGT AATGCTAACG 13320

CTAACACAAG AACCAAAGGG AATGCTGAGT TGATTTTGCA AATTGATGCC TATCTTCCTG 13380

ATACTTATAT TTCTGATCAA CGACATAAGA TTGAAATTTA CAAGAAAATT CGTCAAATTG 13440

ACAACCGTGT CAATTATGAA GAGTTACAAG AGGAGTTGAT AGACCGTTTT GGAGAATACC 13500

CAGATGTAGT AGCCTATCTG TTAGAGATTG GTTTGGTCAA ATCATACTTG GACAAGGTCT 13560

TTGTTCAACG TGTGGAAAGA AAAGATAATA AAATTACAAT TCAATTTGAA AAAGTCACTC 13620

AACGACTGTT TTTAGCTCAA GATTATTTTA AAGCTTTATC CGTAACGAAC TTAAAAGCAG 13680

GCATCGCTGA GAATAAGGGA TTAATGGAGC TTGTATTTGA TGTCCAAAAT AAGAAAGATT 13740

ATGAAATTTT AGAAGGTTTG CTGATTTTTG GAGAAAGTTT ATTAGAGATA AAAGAGTCTA 13800

AGGAAGAAAA TTCCATTTGA TATTTTTCTT CTATAAAATA GATAAAAATG GTACAATAAT 13860

AAATTGAGGT AATAAGGATG AGATTAGATA AATATTTAAA AGTATCGCGA ATTATCAAGC 13920

GTCGTACAGT CGCAAAGGAA GTAGCAGATA AAGGTAGAAT CAAGGTTAAT GGAATCTTGG 13980

CCAAAAGTTC AACGGACTTG AAAGTTAATG ACCAAGTTGA AATTCGCTTT GGCAATAAGT 14040

TGCTGCTTGT AAAAGTACTA GAGATGAAAG ATAGTACAAA AAAAGAAGAT GCAGCAGGAA 14100

TGTATGAAAT TATCAGTGAA ACACGGGTAG AAGAAAATGT CTAAAAATAT TGTACAATTG 14160

AATAATTCTT TTATTCAAAA TGAATACCAA CGTCGTCGCT ACCTGATGAA AGAACGACAA 14220

AAACGGAATC GTTTTATGGG AGGGGTATTG ATTTTGATTA TGCTATTATT TATCTTGCCA 14280

ACTTTTAATT TAGCGCAGAG TTATCAGCAA TTACTCCAAA GACGTCAGCA ATTAGCAGAC 14340

TTGCAAACTC AGTATCAAAC TTTGAGTGAT GAAAAGGATA AGGAGACAGC ATTTGCTACC 14400

AAGTTGAAAG ATGAAGATTA TGCTGCTAAA TATACACGAG CGAAGTACTA TTATTCTAAG 14460

TCGAGGGAAA AAGTTTATAC GATTCCTGAC TTGCTTCAAA GGTGATAAAA TGGAAAATTT 14520

ATTAGACGTA ATAGAGCAAT TTTTGAGTTT GTCAGATGAA AAGCTGGAAG AATTGGCTGA 14580

TAAAAATCAA TTATTGCGTT TACAAGAAGA AAAGGAAAGG AAGAATGCGT AAATTCTTAA 14640

TTATTTTGTT GCTACCAAGT TTTTTGACCA TTTCAAAAGT CGTTAGCACA GAAAAAGAAG 14700

TCGTCTATAC TTCGAAAGAA ATTTATTACC TTTCACAATC TGACTTTGGT ATTTATTTTA 14760

GAGAAAAATT AAGTTCTCCC ATGGTTTATG GAGAGGTTCC TGTTTATGCG AATGAAGATT 14820

TAGTAGTGGA ATCTGGGAAA TTGACTCCCA AAACAAGTTT TCAAATAACC GAGTGGCGCT 14880

TAAATAAACA AGGAATTCCA GTATTTAAGC TATCAAATCA TCAATTTATA GCTGCGGACA 14940

AACGATTTTT ATATGATCAA TCAGAGGTAA CTCCAACAAT AAAAAAAGTA TGGTTAGAAT 15000

CTGACTTTAA ACTGTACAAT AGTCCTTATG ATTTAAAAGA AGTGAAATCA TCCTTATCAG 15060

CTTATTCGCA AGTATCAATC GACAAGACCA TGTTTGTAGA AGGAAGAGAA TTTCTACATA 15120

TTGATCAGGC TGGATGGGTA GCTAAAGAAT CAACTTCTGA AGAAGATAAT CGGATGAGTA 15180

AAGTTCAAGA AATGTTATCT GAAAAATATC AGAAAGATTC TTTCTCTATT TATGTTAAGC 15240

AACTGACTAC TGGAAAAGAA GCTGGTATCA ATCAAGATGA AAAGATGTAT GCAGCCAGCG 15300

TTTTGAAACT CTCTTATCTC TATTATACGC AAGAAAAAAT AAATGAGGGT CTTTATCAGT 15360

TAGATACGAC TGTAAAATAC GTATCTGCAG TCAATGATTT TCCAGGTTCT TATAAACCAG 15420

AGGGAAGTGG TAGTCTTCCT AAAAAAGAAG ATAATAAAGA ATATTCTTTA AAGGATTTAA 15480

TTACGAAAGT ATCAAAAGAA TCTGATAATG TAGCTCATAA TCTATTGGGA TATTACATTT 15540

CAAACCAATC TGATGCCACA TTCAAATCCA AGATGTCTGC CATTATGGGA GATGATTGGG 15600

ATCCAAAAGA AAAATTGATT TCTTCTAAGA TGGCCGGGAA GTTTATGGAA GCTATTTATA 15660

ATCAAAATGG ATTTGTGCTA GAGTCTTTGA CTAAAACAGA TTTTGATAGT CAGCGAATTG 15720

CCAAAGGTGT TTCTGTTAAA GTAGCTCATA AAATTGGAGA TGCGGATGAA TTTAAGCATG 15780

ATACGGGTGT TGTCTATGCA GATTCTCCAT TTATTCTTTC TATTTTCACT AAGAATTCTG 15840

ATTATGATAC GATTTCTAAG ATAGCCAAGG ATGTTTATGA GGTTCTAAAA TGAGGGAACC 15900

AGATTTTTTA AATCATTTTC TCAAGAAGGG ATATTTCAAA AAGCATGCTA AGGCGGTTCT 15960

AGCTCTTTCT GGTGGATTAG ATTCCATGTT TCTATTTAAG GTATTGTCTA CTTATCAAAA 16020

AGAGTTAGAG ATTGAATTGA TTCTAGCTCA TGTGAATCAT AAGCAGAGAA TTGAATCAGA 16080

TTGGGAAGAA AAGGAATTAA GGAAGTTGGC TGCTGAAGCA GAGCTTCCTA TTTATATCAG 16140

CAATTTTTCA GGAGAATTTT CAGAAGCGCG TGCACGAAAT TTTCGTTATG ATTTTTTTCA 16200

AGAGGTCATG AAAAAGACAG GTGCGACAGC TTTAGTCACT GCCCACCATG CTGATGATCA 16260

GGTGGAAACG ATTTTTATGC GCTTGATTCG AGGAACTCGC TTGCGCTATC TATCAGGAAT 16320

TAAGGAGAAG CAAGTAGTCG GAGAGATAGA AATCATTCGT CCCTTCTTGC ATTTTCAGAA 16380

AAAAGACTTT CCATCAATTT TTCACTTTGA AGATACATCA AATCAGGAGA ATCATTATTT 16440

TCGAAATCGT ATTCGAAATT CTTACTTACC AGAATTGGAA AAAGAAAATC CTCGATTTAG 16500

GGATGCAATC TTAGGCATTG GCAATGAAAT TTTAGATTAT GATTTGGCAA TAGCTGAATT 16560

ATCTAACAAT ATTAATGTGG AAGATTTACA GCAGTTATTT TCTTACTCTG AGTCTACACA 16620

AAGAGTTTTA CTTCAAACTT ATCTGAATCG TTTTCCAGAT TTGAATCTTA CAAAAGCTCA 16680

GTTTGCTGAA GTTCAGCAGA TTTTAAAATC TAAAAGCCAG TATCGTCATC CGATTAAAAA 16740

TGGCTATGAA TTGATAAAAG AGTACCAACA GTTTCAGATT TGTAAAATCA GTCCGCAGgC 16800

TGATGAAAAG GAAGATGAAC TTGTGTTACA CTATCAAAAT CAGGTAGCTT ATCAAGGATA 16860

TTTATTTTCT TTTGGACTTC CATTAGAAGG TGAATTAATT CAACAAATAC CTGTTTCACG 16920

TGAAACATCC ATACACATTC GTCATCGAAA AACAGGAGAT GTTTTGATTA AAAATGGGCA 16980

TAGAAAAAAA CTCAGACGTT TATTTATTGA TTTGAAAATC CCTATGGAAA AGAGAAACTC 17040

TGCTCTTATT ATTGAGCAAT TTGGTGAAAT TGTCTCAATT TTGGGAATTG CGACCAATAA 17100

TTTGAGTAAA AAAACGAAAA ATGATATAAT GAACACTGTA CTTTATATAG AAAAAATAGA 17160

TAGGTAAAAA ATGTTAGAAA ACGATATTAA AAAAGTCCTC GTTTCACACG ATGAAATTAC 17220

AGAAGCAGCT AAAAAACTAG GTGCTCAATT AACTAAAGAC TATGCAGGAA AAAATCCAAT 17280

CTTAGTTGGG ATTTTAAAAG GATCTATTCC TTTTATGGCT GAATTGGTCA AACATATTGA 17340

TACACATATT GAAATGGACT TCATGATGGT TTCTAGCTAC CATGGTGGAA CAGCAAGTAG 17400

TGGTGTTATC AATATTAAAC AAGATGTGAC TCAAGATATC AAAGGAAGAC ATGTTCTATT 17460

TGTAGAAGAT ATCATTGATA CAGGTCAAAC TTTGAAGAAT TTGCGAGATA TGTTTAAAGA 17520

AAGAGAAGCA GCTTCTGTTA AAATTGCAAC CTTGTTGGAT AAACCAGAAG GACGTGTTGT 17580

AGAAATTGAG GCAGACTATA CTTGCTTTAC TATCCCAAAT GAGTTTGTAG TAGGTTATGG 17640

TTTAGACTAC AAAGAAAATT ATCGTAATCT TCCTTATATT GGAGTATTGA AAGAGGAAGT 17700

GTATTCAAAT TAGAAAGAAT AATCTTTAAT GAAAAAACAA AATAATGGTT TAATTAAAAA 17760

TCCTTTTCTA TGGTTATTAT TTATCTTTTT CCTTGTGACA GGATTCCAGT ATTTCTATTC 17820

TGGGAATAAC TCAGGAGGAA GTCAGCAAAT CAACTATACT GAGTTGGTAC AAGAAATTAC 17880

CGATGGTAAT GTAAAAGAAT TAACTTACCA ACCAAATGGT AGTGTTATCG AAGTTTCTGG 17940

TGTCTATAAA AATCCTAAAA CAAGTAAAGA AGAAACAGGT ATTCAGTTTT TCACGCCATC 18000

TGTTACTAAG GTAGAGAAAT TTACCAGCAC TATTCTTCCT GCAGATACTA CCGTATCAGA 18060

ATTGCAAAAA CTTGCTACTG ACCATAAAGC AGAAGTAACT GTTAAGCATG AAAGTTCAAG 18120

TGGTATATGG ATTAATCTAC TCGTATCCAT TGTGCCATTT GGAATTCTAT TCTTCTTCCT 18180

ATTCTCTATG ATGGGAAATA TGGGAGGAGG CAATGGCCGT AATCCAATGA GTTTTGGACG 18240

TAGTAAGGCT AAAGCAGCAA ATAAAGAAGA TATTAAAGTA AGATTTTCAG ATGTTGCTGG 18300

AGCTGAGGAA GAAAAACAAG AACTAGTTGA AGTTGTTGAG TTCTTAAAAG ATCCAAAACG 18360

ATTCACAAAA CTTGGAGCCC GTATTCCAGC AGGTGTTCTT TTGGAGGGAC CTCCGGGGAC 18420

AGGTAAAACT TTGCTTGCTA AGGCAGTCGC TGGAGAAGCA GGTGTTCCAT TCTTTAGTAT 18480

CTCAGGTTCT GACTTTGTAG AAATGTTTGT CGGAGTTGGA GCTAGTCGTG TTCGCTCTCT 18540

TTTTGAGGAT GCCAAAAAAG CAGCACCAGC TATCATCTTT ATCGATGAAA TTGATGCTGT 18600

TGGACGTCAA CGTGGAGTCG GTCTCGGCGG AGGTAATGAC GAACGTGAAC AAACCTTGAA 18660

CCAACTTTTG ATTGAGATGG ATGGTTTTGA GGGAAATGAA GGGATTATCG TCATCGCTGC 18720

GACAAACCGT TCAGATGTAC TTGACCCTGC CCTTTTGCGT CCAGGACGTT TTGATAGAAA 18780

AGTATTGGTT GGTCGTCCTG ATGTTAAAGG TCGTGAAGCA ATCTTGAAAG TTCACGCTAA 18840

GAATAAGCCT TTAGCAGAAG ATGTTGATTT GAAATTAGTG GCTCAACAAA CTCCAGGCTT 18900

TGTTGGTGCT GATTTAGAGA ATGTCTTGAA TGAAGCAGCT TTAGTTGCTG CTCGTCGCAA 18960

TAAATCGATA ATTGATGCTT CAGATATTGA TGAAGCAGAA GATAGAGTTA TTGCTGGACC 19020

TTCTAAGAAA GATAAGACAG TTTCACAAAA AGAACGAGAA TTGGTTGCTT ACCATGAGGC 19080

AGGACATACC ATTGTTGGTC TAGTCTTGTC GAATGCTCGC GTTGTCCATA AGGTTACAAT 19140

TGTACCACGC GGCCGTGCAG GCGGATACAT GATTGCACTT CCTAAAGAGG ATCAAATGCT 19200

TCTATCTAAA GAAGATATGA AAGAGCAATT GGCTGGCTTA ATGGGTGGAC GTGTAGCTGA 19260

AGAAATTATC TTTAATGTCC AAACCACAGG AGCTTCAAAC GACTTTGAAC AAGCGACACA 19320

AATGGCACGT GCAATGGTTA CAGAGTACGG TATGAGTGAA AAACTTGGCC CAGTACAATA 19380

TGAAGGAAAC CATGCTATGC TTGGTGCACA GAGTCCTCAA AAATCAATTT CAGAACAAAC 19440

AGCTTATGAA ATTGATGAAG AGGTTCGTTC ATTATTAAAT GAGGCACGAA ATAAAGCTGC 19500

TGAAATTATT CAGTCAAATC GTGAAACTCA CAAGTTAATT GCAGAAGCAT TATTGAAATA 19560

CGAAACATTG GATAGTACAC AAATTAAAGC TCTTTACGAA ACAGGAAAGA TGCCTGAAGC 19620

AGTAGAAGAG GAATCTCATG CACTATCCTA TGATGAAGTA AAGTCAAAAA TGAATGACGA 19680

AAAATAACCC TGAGAGAGGC TGGAGCCTCT CTTTTTTGTG CAGTTTAGGA GCTAAAGGGA 19740

ACAGAATGGA GAAAATGGAA CAAATGTGTT TTCTAATCTG TTAGACTGTA TCTAGAAAGG 19800

GGAAAATTAT GATTAAAGAA TTGTATGAAG AAGTCCAAGG GACTGTGTAT AAGTGTAGAA 19860

ATGAATATTA CCTTCATTTA TGGGAATTGT CGGATTGGGA GCAAGAAGGC ATGCTCTGCT 19920

TACATGAATT GATTAGTAGA GAAGAAGGAC TGGTAGACGA TATTCCACGT TTAAGGAAAT 19980

ATTTCAAGAC CAAGTTTCGA AATCGAATTT TAGACTATAT CCGTAAACAG GAAAGTCAGA 20040

AGCGTAGATA CGATAAAGAA CCCTATGAAG AAGTGGGTGA GATCAGTCAT CGTATAAGTG 20100

AGGGGGGTCT CTGGCTAGAT GATTATTATC TCTTTCATGA AACACTAAGA GATTATAGAA 20160

ACAAACAAAG TAAAGAGAAA CAAGAAGAAC TAGAACGCGT CTTAAGCAAT GAACGATTTC 20220

GAGGGCGTCA AAGAGTATTA AGAGACTTAC GCATTGTGTT TAAGGAGTTT ACTATCCGTA 20280

CCCACTAGTA AGTCATGCAA AAAAAATGAA AAAAATTAGA AAAAGTAGTT GACAAAGTTT 20340

GAAAAGGCTG TATAATAGTA AGAGTTGAAA ATAACAACTC AGGTCCGTTG GTCAAGGGGT 20400

TAAGACACCG CCTTTTCACG GCGGTAACAC GGGTTCGAAT CCCGTACGGA CTATGGTATG 20460

TTGCGTCAGG ACCACTTGAT GAAAAAAAGT TTAAAAAAAC TTAAAAATCT TCAAAAAAGT 20520

GTTGACAAGC GAAAGCAGTT GTGATATACT AATATAGTTG TCGCTTGAGA GAAGCAAGTG 20580

ACAAAGACCT TTGAAAACTG AACAAGACGA ACCAATGTGC AGGGCGCTAC AACGTAAGTT 20640

GTAGTACTGA ACAATGAAAA AAACAATAAA TCTGTCAGTG ACAGAAATGA GTAAGAACTC 20700

AAACTTTTTA ATGAGAGTTT GATCCTGGCT CAGGACGAAC GCTGGCGGCG TGCCTAATAC 20760

ATGCAAGTAG AACGCTGAAG GAGGAGCTTG CTTCTCTGGA TGAGTTGCGA ACGGGTGAGT 20820

AACGCGTAGG TAACCTGCCT GGTAGCGGGG GATAACTATT GGAAACGATA GCTAATACCG 20880

CATAAGAGTA GATGTTGCAT GACATTTGCT TAAAAGGTGC ACTTGCATCA CTACCAGATG 20940

GACCTGCGTT GTATTAGCTA GTTGGTGGGG TAACGGCTCA CCAAGGCGAC GATACATAGC 21000

CGACCTGAGA GGGTGATCGG CCACACTGGG ACTGAGACAC GGCCCAGACT CCTACGGGAG 21060

GCAGCAGTAG GGAATCTTCG GCAATGGACG GAAGTCTGAC CGAGCAACGC CGCGTGAGTG 21120

AAGAAGGTTT TCGGATCGTA AAGCTCTGTT GTAAGAGAAG AACGAGTGTG AGAGTGGAAA 21180

GTTCACACTG TGACGGTATC TTACCAGAAA GGGACGGCTA ACTACGTGCC AGCAGCCGCG 21240

GTAATACGTA GGTCCCGAGC GTTGTCCGGA TTTATTGGGC GTAAAGCGAG CGCAGGCGGT 21300

TAGATAAGTC TGAAGTTAAA GGCTGTGGCT TAACCATA 21338

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6273 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

TGTTTTTAAA GAGCCGTGTC TGGATAGACT TTCGGACGCA ACGCTCTATT AGATAATGAA 60

CTGCCTATAC ACAAGATTTC TAACCTTAGT CGACATGAGC TGAAACCTCT TATTTGTTAA 120

GTAGTTCACA AAATATTATA CACCTATTTT ATGAATAGTC AACTGTCTTT ACAGTAAAAT 180

TTTAGAAAAT CATGAAAATT TTCTCTTTCT TTCCATTTTA AGTGACATTC AGTCATTCTC 240

ACATCAAAAA AGCCCAGACG AAATTGTCTG AGCATTCTTT TATCTAGTCG TTTAAGGAAG 300

TTGAGTTCAG TATGTTTAAA GTCTCTGTCC CATCATTTCT TCAACAAACC TTGTTCTTGG 360

AGAAACTCCT TGGCTACTTG CTTTGCTGAC TTGCCTTCAA CACCGACTTG GTAGTTGAGC 420

TGGCTCATCT GGCTTTCTGT AATCTTACCA GCCAATGTAT TAAGAACTCT TTCCAACTCT 480

GGGTGTTTCT TGAGAAGAGC TTCTTTCATG AGTGGAGCCC CTTGATAAGG TGGGAAGAGT 540

TGCTTGTCAT CTTCCAAGAC CTGTAAATCA TAACGCTCCA ATTCCGCATC AGTCGAATAG 600

GCATCCGTGA TTTGAATATC CCCTGACTGA ATAGCCTGAT AGCGAAGGGC TGGCTCAATG 660

GTCGCTACAT TGAGATTGAG ACCATACATT GATTGCAAGC CCTTATTTCC ATCTTCACGG 720

TCGTTAAACT CGAGTGTAAA ACCTGCCTTC AACTGCCCTT CCACTTTTTT CAAGTCTGAA 780

ATGGTCTTCA AGCCATATTC TTGAGCAATC TTTTTCGGAA CAGCTACAGC ATAGGTGTTT 840

TGATAAGACA TGGGTTTGAG ATAGGCTAGA TGATCCTGCT TAGCAATGCC ATCACGCGCC 900

ACCTGATAAA CCTGTTCTGG TTCATGACTC ACCTTGGGTG ATGGTTGAAG CAAACTTTCA 960

GTCACCGTAC CAGTAAATTC AGGATAGATG TCAATATCGC CTTTTTTCAG AGCTTCATAA 1020

AGGAAGCTTG TCTTCCCAAA ATTCGGTTTA ACAGTCGCAG TCATGCTGGT ATTTTCTTCA 1080

ATCAGCAACT TATACATATT GGCCAAAATT TCTGGTTCTG GACCTATTTT CCCAGCAATA 1140

ACCAAGTTTT CCTTCTCTTT TTGAACCAAA AGAGCTGGAC TATAAGACAG ACCCAGTAAT 1200

AAAGCCACCA AGGCAAAACC TGAGAAAATC GTCCGTAATT TTGCTTTTTC CATCACTTTT 1260

AGTAGGAAGT TAAAGGCAAT GGCTAGCACT GCAGAAGAAA GTGCCCCAAT CAAAATCAAA 1320

CTGGCATTAT TACGGTCAAT TCCCAAAAGA ATAAAGGAAC CTAGTCCCCC TGCACCAATC 1380

AAGGCCGCCA AGGTTGCCGT ACCGATAATC AAAACAGCTG CCGTCCGAAT CCCAGACATG 1440

ATAACAGGCA TGGCGAGTGG AATTTCAAAT TTCTTGAGAC GTTCCCATCT GGTCATCCCA 1500

AAGGCAATCC CAGCCTCTTG CAGGTTCGGA TCAATTCCCT TCAGCCCAGT GATAGTATTT 1560

TGCAAAATAG GGAAAATCGC ATAAATCACT AGAGCTGTCA AAGCCGGCAA GGTCCCAATT 1620

CCCATCAAAG GGATAAAGAG CCCCAACAAG GCCAGAGACG GGATGGTCTG GAAAATACCT 1680

GCAATCTGCA AGACCCAGTC GGCCAGCTTC TCATGATAGC GAAGAAAAAC AGCCAAGGGA 1740

ATCGCAAGCA AAATAGCTAG TAACAAGGTC AAAAGCGACA ACTGCAAATG TTGAGATAGA 1800

GCTGTCAACC AATCACTAAA ACGATCCTGA AAAGTTGCAA TTAAATTAGT CATGAACACT 1860

ACCTCCAAAC AAGTCTGCTA CAAAGTCTGT TGCAGGCGCT TTTAAAATTG TCTCGGGATT 1920

CGCTACCTGG CGAATTTCTC CATCCTGCAA GACAGCAATA CGGTCCGCCA ACTTCAAGGC 1980

TTCATCCGTA TCATGGGTTA CAAAAATCGT TGTCATCCCA AACTCTTTAT GCAATTCTTT 2040

TGTCAGAACC TGCAACTGTT TTCTCGAAAT AGCATCCAAG GCCGAAAAGG GTTCATCCAT 2100

GAGGAAAATC TTGGGCTGAC CAATCATAGC TCGGACAATA CCGACCCGTT GCTGTTCTCC 2160

ACCAGATAAT TCACTAGGTA AGCGATGCCC ATACTCGGCT ACTGGTAAAC CAACCTTAGC 2220

CAAAAGCTCT TCTGTTTTCT TCGTAATTTC TTCCTTGCTC CACCCCTTCA TTTCAGGAAT 2280

GAGAGCAATA TTTTCCGCAA CTGTTAGATT TGGAAAAAGA GCAATAGCCT GTAAAACATA 2340

ACCAGTAGAA AGACGAAGTT CACGCTCATC ATAGTCTTTG ATGCGCTTCC CATCCATATA 2400

AATATTTCCA TCAGTTGGTT CCAAAAGACG GTTAATCATC TTGAGCATGG TCGTCTTACC 2460

TGACCCAGAA GGCCCTACTA AAACCATAAA TTCCCCATCC TCAATCTGTA AGTTGACATC 2520

TCTCAAGACA TCCTTTTCTG TGTAGCGCAG TGCTACATTT TTGTATTCAA TCATTCTTTG 2580

TCCTCAATTT AAAACTTCCC TCGATTGGTC AAGTCTTCTA CCTTAGGCAT AACTTCCTTA 2640

TTATCCCAAT GCTCCACAAT TTTCCCGTTC TCTAAACGGA AGATATCGTA CTGGGCATAA 2700

GCAACGCCAT CAATCTGAGT CTGACCATAG CTAACCACAT AGTTTCCTTG TCCTAAGAGT 2760

TGGAAAACAA AGTCAAAAGT GACACTATAT TCAGCCACAT AGTTTTTATA AGCAGCACTT 2820

CCTTGTCCAA TATCATGATT ATGCTGAATC AAATCGTCTG CCACATAATC ACTCCACTGC 2880

TCTAGCTCCC CATTTTGGAA AATTTCTGTC AAGAAACGGC GAACCAGCTT TTTATTTTCT 2940

GCTTTCTTAT CCAAATCCTT GATTTCAAAA TCTCCAAAAA TTTGATCTAG TTGGTCATTT 3000

TCAGGTGTTC GATAGTAGTC AATGACATCC CAATGCTCAA CAATACAACC ATTCTCATCC 3060

TCACGGAAAG TATCCGTCGT CACCCATTGA GCTTCTCCAC CATTCAGATA TTGATGAACA 3120

TGAACAAAGA CCAGATTGCC ATCCTCAATG GTGCGGACAA TCTTAATCTG ACGCTCTGGA 3180

TGACGCTCAA AGAAATCTGC AAAGAAGGCT GCAAATCCTT CTTTCCCGTC AGGAACACCT 3240

GTCGAATGTT GGATATAGGT ATCCCCTACA GACTGGGCTT GAGCCTCAGC AACTCGTCCG 3300

TCTTGAATGG CATGGATGTA TAGGTTGTGA GCATTTTTCA CTTGTTGTGA CATATTCTAA 3360

ACCTCATTTC CCTTCTCTTT CAGATTCGCC AAAATTCTTT CTTGAAAACC TTCAAATTGG 3420

TGAATTTCTT CCTCTGAAAA TCCTTTGTAA AAGATAGTAT CCAATTTCTG ACTGACACGA 3480

TGCCCCACTT CTTTCTGGGA CTTGCCTAAC TCCGTTAAAA CTAAATACTT CTTACGCTTG 3540

TCTTTTCCAC ACGGACTAAC AATTACAAGC TTTTGTTCCT CTAGCTTTTT TATCATAGTC 3600

GTCAGCGTAT TATTCGCAAG TCCAGTCGCA AGCGCGATAT CTGTCGCAGT TGCGCAGCCA 3660

GTTTCACTAT TCCATAAAAC CGCTAAAATC TTGCCCTGTT CACCCCTATA AAGAGCCTCA 3720

GGATCTTGAC TCAGTAACTT TTGAAAAATC CGCCCATTCA ACAAACGAAT ATGATGGGCT 3780

AGCAAATGAC CATCTTTCAT AACACCTCCA ATTTATTTCG ATATCGAAAT GAATAAAACA 3840

ATTGTAACAC TCATCGTTCT AACTGTCAAC TATTTCGATT TAGAAATAAT TTTTGATAAT 3900

TATCCACACC ACCATACTCC GGCTCAACTA ACTTTTAACG AGAGTTTCTA AACTCCTTCG 3960

TCCTCCAGTC TACAAAAGCC TTCCATTCGT ACTATCCTAT ATTTTATGAG GGGACACATT 4020

TTTCCTATCA GACCATTTAT TTTAAAGATA GAAGTAAATC ATAATTGCTT CCATCTGTTC 4080

TTTTATAGTA TATTGAAGTT AGACTAGAGC ACTGTATCTT CTAAAACATT GATAGAAAGC 4140

GATTTGAATT TCCCAATCAA TTTGTTCGTA TTTATAGCAT TTCGAAACTG GAATAGGACA 4200

CCATGACTGC TAAAAGATTT CTATAAATTC ATTTAATTTC CTCAATCAAT TTGTTCATAT 4260

CTTATTTCAT TCCGCTATAA TTTCACCTTA CCCTATCTTT TTCGTAGCAC CCTTCAAACA 4320

GCCTATCCCC TACCGTTTGA CGATTCCTCA CTTCGCTCCA CTTCCATTAC AGAAGTTTCT 4380

TCACTACTAT GGGCTCGGCT GACTTCTCAT GATTCCTTGT TACTACTATT TGAACGCTCA 4440

CGAGATAGAT CTTACAAAAA ATGCTTTGAT CCACAATGGA ATCAAAGCAT TTTAAAGAGT 4500

TCCTCATACA TAAGCGCAGA AGTCGCAGTT CCTCTGTACT TGGCTTCTTC TCTTTTGACA 4560

AAGCGAGCCA AGTTGAGCAA CTCAGGTGCT GGATGTTTGG GATTTAGGAG CAATTCACGA 4620

TTGACCAGGC CTGAGAGACG AACTGCCTGC AATTGCTCAT TTGTAGTAGG CAGTTTTTTA 4680

GTAGTCTCTA GGAGAGCAGC AACTAAATCT TCACTCAAAT CATGTCGAGC ATGATTGTAA 4740

AGATCTTTTA TAAGGCTTTC TAGGTTTGGT TCTACCATCC CTACCACCTC CCTTATGGTT 4800

TAATAATGTT TAATCAAATC AACCGTTGAA CGATCCAATT TCTTCACCAA GGCTTGTAAG 4860

AAAGCTTGCG CTTCTAGGAA GTCATCCATT GCATAGAGGG TTTGGTGAGA ATGGATATAA 4920

CGAGCGCAGA CACCGATAGT TGTTGATGGG ACACCACCAT TTTTCAGATG AGCTGCACCT 4980

GCATCTGTTC CGCCTTTACC ACAGTAGTAT TGGTACTTGA TACCAGCTTC TTCAGCCGTT 5040

GTCAAAAGGA AATCCTTCAT CCCTGGGAGA AGCAAGTGAC CTGGATCATA GAAACGAATC 5100

AAGGTTCCAT CTCCAATCTT GCCTTGACCA CCGTAGACAT CACCTGCTGG TGAGCAATCA 5160

ACTGCGAGGA AGACTTCTGG GTCAAACTTG GTTGTAGAGG TATGAGCGCC ACGCAGACCA 5220

ACTTCTTCTT GGACGTTAGA ACCCAGATAG AGTTCATTGC CGAGTTTTTG ACCCGATAAA 5280

GCTTCAGCTA GCTCGCTTAC CATGAGGACA CCGTAGCGGT TATCCCAAGC TTTTGAGATG 5340

ATATTTTTTT CATTGGCTGT CAAAATTGCA GAACTATCTG GTACAATGGT ATCACCAGGA 5400

CGGATGCCAA AACTTTCTGC CTCAGCCTTG TCCGCAAAAC CACCATCAAA AACGATATCG 5460

GCAATGGCTG GCATGGTTGG TCCCCCCTTT CCACGAGTCA AATGCGGAGG AACAGAACCT 5520

GAAATCACAG GAATTTCATG ACCATCACGA GTCAAGAGTT TGAAACGTTG GCTGCTAACC 5580

ACCATGGGGT TCCAGCCACC GATTTCTACG ACACGGAAGG TACCATCTGG CTTGATTTCG 5640

CTGACCATAA AACCAACTTC GTCCATATGA GAAGCGACCA AGACGCGCGG TGCATCCACA 5700

GCTTCTGAAT GTTTGATACC AAAAATACCA CCCAAGCCAT CTGTCACCAC TTCATCCACA 5760

TGCGGTGTCA ACTTTTCACG AAGATAAGCA CGGACAGGCG CTTCATGACC TGAGACTGCA 5820

GCAAGTTCTG TTACTTCTTT AATTTTTGAA AATAATGTTG TCATTTCAGT TCCTTCTTTC 5880

TTTCATCCAT TTTACCACTT TTTATAGGAG AAGGATAGTG GGAAGGTGGA TTTCTAAGTT 5940

AGTATCTTAG TCCTGCTCTA TCTTAGAAAA GGATAGTATT CTCTTGCATG TAGTGCAAAA 6000

TCTAGTAAAC ATTCCAAAAT TAACTCGAAT ATTTATTTCC AAACAAAAAA ACAATACACC 6060

ATCAAAGTTG TTTGGATTTT TCATGAAATT TACAGAAAAT AGTTGACTTC CCTTTCTTCT 6120

TTCTTTAAAT ATATAGTTGG TTGAGTTTGG AATAGTACGC TGTAGCTGCT AAAACATTTC 6180

TAGAAATTAA TTTGACTTTC CTAATAGAGT TGTTCATATC TTATTTCAAT TTACTATAGT 6240

ACAAAACTAG AAAAGGAAAA AATCATGACC AGG 6273

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28171 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

ACAACCTTTT TCAAAAACTC ACCTTGGTAC GGAGATGTTT TGCTTTCTGC TATTATTTTC 60

GGTTATATTC ATATCAATTT TGCTTTAACT CCTCTTGCTT TTTTCATTTA TGCTAGTGGA 120

GGTCTTATTT TAGCTCTATT GTATCGCATG ACTAAAAATC TCTACTATCC AATACTAGTT 180

CATATTCTCA TTAATATCAC TGCCTTCTGG GATGTGTGGT TGCTCCTATT TTCAGGAAGT 240

TAGCTTACTA AAATAATGTC GGAACTTTCC GGCATTTTCT TTTTTCACAA ATAGTCAACG 300

TTTTTCTTTT CGATATTGTA GTGGTGTGTA TCCAGTTATT TTTTTGAATT GATTTTGAAA 360

ATAAGGTTGA CTTGAGAAAG GCAGATAGTG AAGATAGTTA AGAAGAATAG GATGTTCTTT 420

TTTCCTTTTT GGAAAACTTC TAAAATATGG TATAATGAAA AGATAAAGAA GTTGGGGGTA 480

GAAGATGAAC ATTCAACAAT TACGCTATGT TGTGGCTATT GCCAATAGTG GTACTTTTCG 540

TGAAGCTGCT GAAAAGATGT ATGTTAGTCA GCCGAGTCTG TCTATTTCTG TTCGTGATTT 600

GGAAAAAGAG TTGGGCTTTA AGATTTTCCG TCGGACCAGC TCAGGGACTT TCTTGACCCG 660

TCGTGGGATG GAATTTTATG AAAAATCGCA AGAATTGGTT AAAGGATTTG ATATTTTTCA 720

AAATCAGTAT GCCAATCCTG AAGAAGAAAA AGATGAATTT TCTGTTGCTA GCCAGCACTA 780

TGACTTCTTG CCACCAACTA TTACGGCCTT TTCAGAGCGC TATCCTGACT ATAAGAACTT 840

CCGTATTTTT GAATCAACTA CTGTTCAAAT ATTAGATGAA GTGGCGCAAG GGCATAGTGA 900

GATTGGGATT ATCTACCTCA ACAATCAAAA TAAAAAGGGG ATTATGCAAC GGGTTGAAAA 960

ATTAGGTCTG GAGGTCATCG AATTGATTCC TTTCCATACC CATATTTATC TCCGTGAGGG 1020

TCATCCTTTA GCCCAGAAAG AGGAATTAGT CATGGAGGAT TTAGCGGATT TACCAACGGT 1080

TCGTTTCACT CAAGAGAAAG ACGAGTACCT TTATTATTCA GAGAACTTTG TCGATACCAG 1140

CGCTAGCTCA CAGATGTTTA ATGTGACAGA CCGTGCCACC TTGAATGGTA TTTTGGAGCG 1200

GACGGACGCC TATGCGACAG GTTCTGGATT TTTAGATAGT GACAGTGTTA ATGGCATTAC 1260

AGTTATTCGT CTCAAGGATA ACCTAGATAA CCGCATGGTC TATGTTAAAC GTGAAGAAGT 1320

GGAGCTTAGT CAAGCTGGGA CTCTCTTCGT AGAAGTCATG CAAGAATATT TTGATCAAAA 1380

GAGGAAATCA TGAAAAAAAG AGCAATAGTG GCAGTCATTG TACTGCTTTT GATTGGGCTG 1440

GATCAGTTGG TCAAATCCTA TATCGTCCAG CAGATTCCAC TGGGTGAAGT GCGCTCCTGG 1500

ATCCCCAATT TCGTTAGCTT GACCTACCTG CAAAATCGAG GTGCAGCCTT TTCTATCTTA 1560

CAAGATCAGC AGCTGTTATT CGCTGTCATT ACTCTGGTTG TCGTGATAGG TGCCATTTGG 1620

TATTTACATA AACACATGGA GGACTCATTC TGGATGGTCT TGGGTTTGAC TCTAATAATC 1680

GCGGGTGGTC TTGGAAACTT TATTGACAGG GTCAGTCAGG GCTTTGTTGT GGATATGTTC 1740

CACCTTGACT TTATCAACTT TGCAATTTTC AATGTGGCAG ATAGCTATCT GACGGTTGGA 1800

GTGATTATTT TATTGATTGC AATGCTAAAA GAGGAAATAA ATGGAAATTA AAATTGAAAC 1860

TGGTGGTCTG CGTTTGGATA AGGCTTTGTC AGATTTGTCA GAATTATCAC GTAGTCTCGC 1920

GAATGAACAA ATTAAATCAG GCCAGGTCTT GGTCAATGGT CAAGTCAAGA AAGCTAAATA 1980

CACAGTCCAA GAGGGTGATG TCGTCACTTA CCATGTGCCA GAACCAGAGG TATTAGAGTA 2040

TGTGGCTGAG GATCTTCCGC TAGAAATAGT CTACCAAGAT GAGGATGTGG CTGTCGTTAA 2100

CAAACCTCAG GGAATGGTTG TGCACCCGAG TGCTGGTCAT ACCAGTGGAA CCCTAGTAAA 2160

TGCCCTCATG TATCATATTA AGGACTTGTC GGGTATCAAT GGGGTTCTGC GTCCAGGGAT 2220

TGTTCACCGT ATTGATAAGG ATACGTCAGG TCTTCTCATG ATTGCTAAAA ACGATGATGC 2280

GCATCTAGCA CTTGCCCAAG AACTCAAGGA TAAAAAGTCT CTCCGCAAAT ATTGGGCGAT 2340

TGTTCATGGA AATCTACCTA ATGATCGTGG TGTAATTGAA GCGCCGATTG GCCGGAGTGA 2400

AAAAGACCGT AAGAAACAGG CTGTAACTGC TAAAGGGAAG CCTGCAGTGA CGCGTTTTCA 2460

CGTCTTGGAA CGCTTTGGCG ATTATAGCTT AGTAGAGTTG CAACTGGAGA CAGGGCGCAC 2520

TCATCAAATC CGTGTCCACA TGGCTTATAT CGGCCATCCA GTCGCTGGTG ATGAGGTCTA 2580

TGGTCCTCGC AAGACTTTGA AAGGACATGG ACAATTTCTT CATGCCAAGA CTTTAGGTTT 2640

TACTCATCCG AGAACAGGTA AGACCTTGGA ATTTAAAGCA GATATCCCAG AGATTTTTAA 2700

GGAAACCTTG GAGAGATTGA GAAAGTAAGA ATGAAAAAGA AATTAACTAG TTTAGCACTT 2760

GTAGGCGCTT TTTTAGGTTT GTCATGGTAT GGGAATGTTC AGGCTCAAGA AAGTTCAGGA 2820

AATAAAATCC ACTTTATCAA TGTTCAAGAA GGTGGCAGTG ATGCGATTAT TCTTGAAAGC 2880

AATGGACATT TTGCCATGGT GGATACAGGA GAAGATTATG ATTTCCCAGA TGGAAGTGAT 2940

TCTCGCTATC CATGGAGAGA AGGAATTGAA ACGTCTTATA AGCATGTTCT AACAGACCGT 3000

GTCTTTCGTC GTTTGAAGGA ATTGGGTGTC CAAAAACTTG ATTTTATTTT GGTGACCCAT 3060

ACCCACAGTG ATCATATTGG AAATGTTGAT GAATTACTGT CTACCTATCC AGTTGACCGA 3120

GTCTATCTTA AGAAATATAG TGATAGTCGT ATTACTAATT CTGAACGTCT ATGGGATAAT 3180

CTGTATGGCT ATGATAAGGT TTTACAGACT GCTGCAGAAA AAGGTGTTTC AGTTATTCAA 3240

AATATCACAC AAGGGGATGC TCATTTTCAG TTTGGGGACA TGGATATTCA GCTCTATAAT 3300

TATGAAAATG AAACTGATTC ATCGGGTGAA TTAAAGAAAA TTTGGGATGA CAATTCCAAT 3360

TCCTTGATTA GCGTGGTGAA AGTCAATGGC AAGAAAATTT ACCTTGGGGG CGATTTAGAT 3420

AATGTTCATG GAGCAGAAGA CAAGTATGGT CCTCTCATTG GAAAAGTTGA TTTGATGAAG 3480

TTTAATCATC ACCATGATAC CAACAAATCA AATACCAAGG ATTTCATTAA AAATTTGAGT 3540

CCGAGTTTGA TTGTTCAAAC TTCGGATAGT CTACCTTGGA AAAATGGTGT TGATAGTGAG 3600

TATGTTAATT GGCTCAAAGA ACGAGGAATT GAGAGAATCA ACGCAGCCAG CAAAGACTAT 3660

GATGCAACAG TTTTTGATAT TCGAAAAGAC GGTTTTGTCA ATATTTCAAC ATCCTACAAG 3720

CCGATTCCAA GTTTTCAAGC TGGTTGGCAT AAGAGTGCAT ATGGGAACTG GTGGTATCAA 3780

GCGCCTGATT CTACAGGAGA GTATGCTGTC GGTTGGAATG AAATCGAAGG TGAATGGTAT 3840

TACTTTAACC AAACGGGTAT CTTGTTACAG AATCAATGGA AAAAATGGAA CAATCATTGG 3900

TTCTATTTGA CAGACTCTGG TGCTTCTGCT AAAAATTGGA AGAAAATCGC TGGAATCTGG 3960

TATTATTTTA ACAAAGAAAA CCAGATGGAA ATTGGTTGGA TTCAAGATAA AGAGCAGTGG 4020

TATTATTTGG ATGTTGATGG TTCTATGAAG ACAGGATGGC TTCAATATAT GGGGCAATGG 4080

TATTACTTTG CTCCATCAGG GGAAATGAAA ATGGGCTGGG TAAAAGATAA AGAAACCTGG 4140

TACTATATGG ATTCTACTGG TGTCATGAAG ACAGGTGAGA TAGAAGTTGC TGGTCAACAT 4200

TATTATCTGG AAGATTCAGG AGCTATGAAG CAAGGCTGGC ATAAAAAGGC AAATGATTGG 4260

TATTTCTACA AGACAGACGG TTCACGAGCT GTGGGTTGGA TCAAGGACAA GGATAAATGG 4320

TACTTCTTGA AAGAAAATGG TCAATTACTT GTGAACGGTA AGACACCAGA AGGTTATACT 4380

GTGGATTCAA GTGGTGCCTG GTTAGTGGAT GTTTCGATCG AGAAATCTGC TACAATTAAA 4440

ACTACAAGTC ATTCAGAAAT AAAAGAATCC AAAGAAGTAG TGAAAAAGGA TCTTGAAAAT 4500

AAAGAAACGA GTCAACATGA AAGTGTTACA AATTTTTCAA CTAGTCAAGA TTTGACATCC 4560

TCAACTTCAC AAAGCTCTGA AACGAGTGTA AACAAATCGG AATCAGAACA GTAGTAGAAA 4620

AGAAGGTTTT AGGGCCTTCT TTTTCCTATC AACTCTTTTC TATTTCCTGT TATTCATGTT 4680

ATAATGGATA AATATGAATA ATCGGAGTGA GACTATGAAA TACAAACGGA TTGTCTTTAA 4740

GGTGGGTACT TCTTCTCTGA CAAATGAGGA TGGAAGTTTA TCACGTAGTA AGGTAAAGGA 4800

TATTACCCAG CAGTTGGCTA TGCTGCACGA GGCTGGTCAT GAGTTGATTT TGGTGTCTTC 4860

AGGTGCCATT GCGGCTGGTT TTGGAGCCTT AGGATTTAAA AAGCGTCCGA CTAAGATTGC 4920

TGATAAACAG GCTTCAGCAG CGGTAGGGCA GGGGCTTTTG TTGGAAGAAT ATACAACCAA 4980

TCTTCTCTTG CGTCAAATCG TTTCTGCACA AATCTTGCTG ACCCAAGATG ACTTTGTGGA 5040

TAAGCGTCGT TATAAAAATG CCCATCAGGC TTTGTCGGTT TTGCTCAACC GTGGGGCAAT 5100

TCCTATCATC AATGAGAATG ATAGTGTCGT TATTGATGAG CTCAAGGTTG GGGACAATGA 5160

CACTCTAAGT GCTCAAGTAG CGGCGATGGT CCAAGCAGAC CTTTTAGTTT TCTTGACAGA 5220

TGTGGACGGT CTCTATACTG GAAATCCTAA TTCAGATCCA AGAGCCAAAC GCTTGGAGAG 5280

AATCGAGACC ATCAATCGTG AGATTATTGA TATGGCTGGT GGAGCTGGTT CGTCAAACGG 5340

AACTGGGGGT ATGTTAACCA AAATCAAGGC TGCAACTATC GCGACGGAAT CAGGAGTTCC 5400

TGTTTATATC TGCTCATCCT TGAAATCAGA TTCCATGATT GAGGCGGCAG AGGAGACCGA 5460

GGATGGTTCT TACTTTGTTG CTCAAGAGAA GGGGCTTCGT ACCCAGAAAC AATGGCTTGC 5520

CTTCTATGCT CAGAGTCAAG GTTCTATTTG GGTTGATAAA GGGGCTGCGG AAGCTCTCTC 5580

TCAATATGGA AAGAGTCTTC TCTTATCTGG TATCGTTGAA GCAGAAGGAG TCTTTTCTTA 5640

CGGTGATATC GTGACAGTAT TTGACAAGGA AAGTGGAAAA TCACTTGGAA AAGGACGCGT 5700

GCAATTTGGA GCATCTGCTT TGGAGGATAT GTTGCGTTCT CAAAAAGCCA AGGGTGTCTT 5760

GATTTACCGT GACGACTGGA TTTCCATTAC TCCTGAAATC CAACTACTTT TTACAGAATT 5820

TTAGAGGTAA ACTATGGTGA GTAGACAAGA ACAATTTGAA CAGGTACAGG CTGTTAAAAA 5880

ATCGATTAAC ACAGCTAGTG AAGAAGTGAA AAACCAAGCC TTGCTAGCCA TGGCTGATCA 5940

CTTAGTGGCT GCTACTGAGG AAATTTTAGC GGCTAATGCC CTCGATATGG CAGCGGCTAA 6000

GGGGAAAATC TCAGATGTGA TGTTGGATCG TCTTTATTTG GATGCAGATC GTATAGAAGC 6060

GATGGCAAGA GGAATTCGTG AAGTGGTTGC CTTACCAGAT CCAATCGGTG AAGTTTTAGA 6120

AACAAGTCAG CTTGAAAATG GTTTGGTTAT CACAAAAAAA CGTGTAGCTA TGGGTGTCAT 6180

CGGTATTATC TATGAAAGCC GTCCAAATGT GACGTCTGAT GCGGCTGCTT TGACTCTTAA 6240

GAGTGGAAAT GCGGTTGTTC TTCGTAGTGG TAAGGATGCC TATCAAACAA CCCATGCCAT 6300

TGTCACAGCC TTGAAGAAGG GCTTGGAGAC GACTACTATT CATCCAAATG TGATTCAACT 6360

GGTGGAGGAT ACTAGCCGTG AAAGTAGTTA TGCTATGATG AAGGCCAAGG GCTATCTAGA 6420

CCTTCTCATT CCTCGTGGAG GAGCTGGCTT GATCAATGCA GTGGTTGAGA ATGCGATTGT 6480

ACCTGTTATC GAGACAGGGA CTGGGATTGT CCATGTCTAT GTGGATAAGG ATGCAGACGA 6540

AGACAAGGCG CTGTCTATCA TCAACAATGC TAAAACCAGT CGTCCTTCTG TTTGTAATGC 6600

CATGGAGGTT CTGCTGGTTC ATGAAAACAA GGCAGCAAGC TTCCTTCCTC GCTTGGAGCA 6660

AGTGTTGGTT GCAGAGCGTA AGGAAGCTGG ACTGGAACCA ATTCAATTCC GCCTAGATAG 6720

CAAAGCAAGC CAGTTTGTTT CAGGTCAAGC AGCTGAGACC CAAGACTTTG ACACCGAGTT 6780

TTTAGACTAT GTCCTTGCTG TTAAGGTTGT GAGCAGTTTA GAAGAAGCGG TTGCGCACAT 6840

TGAATCCCAC AGCACCCATC ATTCGGATGC TATTGTGACG GAAAATGCTG AAGCTGCAGC 6900

ATACTTTACA GATCAAGTGG ACTCTGCAGC GGTGTATGTT AATGCCTCAA CTCGTTTCAC 6960

AGATGGAGGA CAATTTGGTC TTGGTTGTGA AATGGGGATT TCTACTCAGA AATTGCACCC 7020

GCGTGGTCCC ATGGGCTTGA AAGAGTTGAC CAGCTACAAG TATGTGGTTG CCGGTGATGG 7080

GCAGATAAGG GAGTAAGAGA TGAAGATTGG ATTTATCGGT TTGGGGAATA TGGGTGCTAG 7140

CTTGGCAAAA TCTGTCTTGC AGACTAGGAC GTCAGATGAG ATTCTCCTTG CCAATCGTAG 7200

TCAAGCTAAG GTAGATGCTT TCATTGCAGA CTTTGGTGGT CAGGCTTCCA GCAATGAAGA 7260

AATGTTTGCA GAAGCAGATG TGATTTTTCT AGGAGTTAAG CCTGCTCAGT TTTCTGAACT 7320

GCTTTCTCAA TACCAGACCA TCCTTGAAAA AAGAGAAAGT CTTCTTTTGA TTTCGATGGC 7380

AGCTGGATTG ACCTTAGAAA AACTAGCAAG TCTTATCCCA AGTCAACACC GAATTATTCG 7440

TATGATGCCT AATACCCCTG CTTCTATCGG GCAAGGAGTG ATTAGTTATG CCTTGTCTCC 7500

TAATTGCAGG GCTGAGGACA GTGAGCTCTT TTATCAGCTT TTAGCCAAGG CTGGTCTCTT 7560

GGTTGAACTA GGAGAAAGTT TAATCGATGC AGCGACAGGT CTTGCAGGTT GTGGACCAGC 7620

CTTTGTCTAT CTTTTTATCG AGGCCTTGGC AGATGCAGGT GTTCAGACAG GATTACCACG 7680

AGAAATAGCA TTGAAAATGG CAGCACAAAC TGTGGTAGGA GCTGGGCAAT TGGTCCTTGA 7740

AAGTCAGCAA CATCCTGGAG TATTGAAAGA CCAAGTCTGT AGCCCAGGCG GTTCGACTAT 7800

CGCTGGTGTA GCAAGCCTAG AAGCGCATGC TTTCCGAGGA ACAGTCATGG ATGCAGTTCA 7860

TCAAGCCTAC AAACGAACAC AAGAACTAGG TAAATAAGAG GTAGTTTTGA CTGCCTCTTT 7920

TATGGTGGCT GAAATGAGAA GACACAAAAA GATTGTCACA AACCCCTATT TTTTTGATAG 7980

AATAGAAGTA GTAAAAAAGA AATGAGTTAG ACATGTCAAA AGGATTTTTA GTCTCTCTTG 8040

AGGGACCAGA GGGAGCAGGC AAGACCAGTG TTTTAGAGGC TCTGCTACCA ATTTTAGAGG 8100

AAAAAGGAGT AGAGGTGTTG ACGACCCGTG AACCTGGCGG AGTCTTGATT GGGGAGAAGA 8160

TTCGGGAAGT GATTTTGGAT CCAAGTCATA CTCAGATGGA TGCTAAAACA GAGCTACTTC 8220

TCTATATTGC CAGTCGCAGA CAGCATTTGG TGGAAAAAGT TCTTCCAGCC CTTGAAGCTG 8280

GCAAGTTGGT CATCATGGAT CGTTTTATCG ATAGTTCTGT TGCCTATCAG GGATTTGGTC 8340

GTGGCTTAGA TATTGAAGCC ATTGACTGGC TCAATCAGTT TGCGACAGAT GGCCTCAAAC 8400

CCGATTTGAC ACTCTATTTT GACATCGAGG TGGAAGAAGG GCTGGCTCGT ATTGCTGCTA 8460

ATAGTGACCG CGAGGTTAAT CGTTTGGATT TGGAAGGGTT GGACTTGCAT AAAAAAGTTC 8520

GTCAAGGCTA CCTTTCTCTT CTGGATAAAG AGGGAAATCG CATTGTCAAG ATTGATGCTA 8580

GTCTCCCTTT GGAGCAAGTT GTGGAAACTA CCAAGGCTGT CTTGTTTGAC GGAATGGGCT 8640

TGGCCAAATG AAACAAGATC AACTAAAGGC TTGGCAACCA GCTCAGTTTG ACCGTTTTGT 8700

CCGTATCTTA GAACAAGACC AGCTCAATCA CGCCTATCTC TTTTCAGGTT TCTTTGAAAG 8760

CTTGGAAATG GCGCAATTTT TAGCTAAGAG CCTCTTTTGT ACGGATAAAG TTGGCGTCTT 8820

ACCATGTGAG AAATGCCGAA GTTGCAAGCT GATTGAACAG GGAGAATTTC CCGATGTCAC 8880

CTTGATTAAA CCAGTTAATC AGGTCATTAA GACGGAACGC ATTCGAGAAT TGGTGGGTCA 8940

GTTTTCTCAA GCAGGGATTG AAAGCCAGCA ACAGGTCTTT ATCATCGAGC AAGCGGATAA 9000

AATGCATCCC AACGCAGCCA ATTCTCTGCT CAAGGTCATC GAAGAACCCC AGAGTGAAGT 9060

TTATATTTTC TTCTTGACTA GCGATGAGGA AAAGATGTTA CCGACAATCC GAAGTCGGAC 9120

TCAGATCTTC CACTTTAAAA AGCAAGAAGA AAAACTTATC TTACTCTTAG AACAAATGGG 9180

ACTTGTTAAG AAAAAAGCGA CTCTTTTAGC TAAGTTTAGT CAATCGCGAG CTGAAGCAGA 9240

AAAGTTGGCT AATCAGGCAA GTTTTTGGAC CTTGGTCGAT GAAAGTGAAC GCCTGCTGAC 9300

TTGGTTAGTA GCTAAGAAAA AAGAAAGTTA TCTACAGGTT GCCAAATTAG CCAACTTGGC 9360

AGATGATAAG GAAAAACAGG ATCAGGTTTT ACGGATTCTT GAAGTTCTCT GTGGGCAGGA 9420

CCTCTTGCAG GTAAGAGTAA GAGTGATTCT ACAAGATTTA CTAGAAGCTA GAAAAATGTG 9480

GCAAGCTAAT GTCAGCTTTC AAAATGCCAT GGAATATCTG GTCTTGAAAG AAATATAAAC 9540

TCAAAAATGA ATGATAAAGA AAGGAAAGGG CTGTTTTATG GACAAAAAAG AATTATTTGA 9600

CGCGCTGGAT GATTTTTCCC AACAATTATT GGTAACCTTA GCCGATGTGG AAGCCATCAA 9660

GAAAAATCTC AAGAGCCTGG TAGAGGAAAA TACAGCTCTT CGCTTGGAAA ATAGTAAGTT 9720

GCGAGAACGC TTGGGTGAGG TGGAAGCAGA TGCTCCTGTC AAGGCCAAGC ATGTTCGTGA 9780

AAGTGTCCGT CGCATTTACC GTGATGGATT TCACGTATGT AATGATTTTT ATGGACAACG 9840

TCGAGAGCAG GACGAGGAAT GTATGTTTTG TGACGAGTTG CTATACAGGG AGTAGGCATG 9900

CAGATTCAAA AAAGTTTTAA GGGGCAGTCT CCCTATGGCA AGCTGTATCT AGTGGCAACG 9960

CCGATTGGCA ATCTAGATGA TATGACTTTT CGTGCTATCC AGACCTTGAA AGAAGTGGAC 10020

TGGATTGCTG CTGAGGATAC GCGCAATACA GGGCTTTTGC TCAAGCATTT TGACATTTCC 10080

ACCAAGCAGA TCAGTTTTCA TGAGCACAAT GCCAAGGAAA AAATTCCTGA TTTGATTGGT 10140

TTCTTGAAAG CAGGGCAAAG TATTGCTCAG GTCTCTGATG CCGGTTTGCC TAGCATTTCA 10200

GACCCTGGTC ATGATTTAGT TAAGGCAGCT ATTGAGGAAG AAATTGCAGT TGTGACAGTT 10260

CCAGGTGCCT CTGCAGGAAT TTCTGCCTTG ATTGCCAGTG GTTTAGCGCC ACAGCCACAT 10320

ATCTTTTACG GTTTTTTACC GAGAAAATCA GGTCAGCAGA AGCAATTTTT TGGCTTGAAA 10380

AAAGATTATC CTGAAACACA GATTTTTTAT GAATGAGCTC ATGGTGTAGC AGACACGTTG 10440

GAAAATATGT TAGAAGTCTA CGGTGACCGC TCCGTTGTCT TGGTCAGGGA ATTGACCAAA 10500

ATCTATGAAG AATACCAACG AGGTACTATC TCTGAGTTAT TAGAAAGCAT TGCTGAAACG 10560

CCACTCAAGG GCGAATGTCT TCTCATTGTT GAGGGTGCCA GTCAGGGTGT GGAGGAAAAG 10620

GACGAGGAAG ACTTGTTCGT AGAAATTCAA ACCCGCATCC AGCAAGGTGT GAAGAAAAAC 10680

CAAGCTATCA AGGAAGTCGC TAAGATTTAC CAGTGGAATA AAAGTCAGCT CTACGCTGCC 10740

TACCACGACT GGGAAGAAAA ACAATAAAGG GAGACAGGAT GTAATAATTC TGTCTGTTTC 10800

TGTTTAACTT AATTAGTGAT GATAATATAA AGATGTATCA CTTGGTATAG AAGCTTTGGT 10860

ATTAAGTTTT TTATTAAGCC CATACGGAAT ACCGATGGTT GGAGCAGCAG TTATAGCGTT 10920

CTTAGAAGGT ATAAATAGAA AAATAAGGTC ATTTTAAATC AAAGGATTGA TAAATCAGAA 10980

AGAAGGTGAT TTTTTGCGAA CATACGAAAA TAAAGAAGAA CTAAAAGCTG AGATAGAGAA 11040

AACATTTGAG AAATATATTT TAGAATTTGA TAATATTCCA GAAAATTTAA AAGATAAGAG 11100

AGCTGATGAA GTTGACAGAA CTCCAGCAGA AAACCTTGCT TATCAGGTTG GTTGGACCAA 11160

CTTGGTTCTT AAATGGGAAG AAGATGAAAG AAAGGGGCTT CAAGTAAAAA CACCATCGGA 11220

TAAATTTAAA TGGAATCAAC TTGGTGAATT ATATCAGTGG TTCACAGATA CCTACGCTCA 11280

TTTATCTCTG CAAGAGTTGA AAGCAAAATT AAATGAAAAT ATTAATTCTA TCTCTGCAAT 11340

GATTGATTCG TTGAGTGAGG AAGAATTATT TGAACCGCAT ATGAGAAAGT GGGCTGATGA 11400

AGCGACTAAA ACAGCGACTT GGGAAGTGTA TAAGTTTATT CATGTAAATA CGGTTGCACC 11460

TTTTGGAACT TTCAGAACTA AAATCAGAAA ATGGAAGAAG ATAGTATTAT AAATTATATT 11520

TTTAACTTTA AAAAATTTCA TAAAAATGGT TACCAAAGGC GATAGAAGAA AAACTATCGT 11580

CTTTTTCTTT GCAAATTTTT AAGAAGGGAG GTGATCTTGC ATGGACTTTG AATATTTTTA 11640

TAACAGAGAA GCGGAAAGAT TTAACTTCTT AAAAGTACCG GAGATATTAG TTGATAGAGA 11700

AGAATTTCGG GGCTTATCAG CAGAAGCAAT TATCCTTTAT TCCATACTTC TTAAACAGAC 11760

AGGAATGTCA TTTAAGAATA ACTGGATAGA CAAGGAAGGC AGAGTATTTA TCTATTTTAC 11820

TGTCGAAGAA ATTATGAAAA GAAGAAATAT CTCAAAGCCA ACTGCCATAA AAACATTAGA 11880

TGAGCTTGAT GTAAAAAAGG AATAGGACTG ATCGAAAGAG TAAGGCTTGG ACTTGGTAAG 11940

CCGAACATCA TTTATGTTAA AGACTTTATG AGTATATTTC AGGTAAAAGA AAATGACTTA 12000

CAGAAGTCAA AAAACTTAAC TTCAGAAGTA AAAGATTTTA ACCTCAGAAG TAAAGAAAAT 12060

GAACTTCAAG AGGTTAAGAA CCTTGACTCT AACTATATAG AGAATAATAA GAGTAAGTAT 12120

AGTAAGAGAG AATATAGTTT TGGTGAAAAC GGACTTGGAA CATTTCAAAA TGTGTTTTTA 12180

GCTGCTGAAG ATATATCGGA TTTACAAATC ATAATGAACT CACAGCTTGA GAATTACATT 12240

AGACTTCCTG CAAAACTAGA ATCCTAGTTC ATGATTGATA ATGCCAGCAA TCAAATTCAT 12300

TCGTAATCCG AAGCGTTTAC GATGATTTCG ATAGATTGTT GAAAACATTT TAAACGTTTT 12360

TACTTTGGCA AAGATGTTCT CAATCTTGCT TCTCTCCTTG GATAGCGCAT GGTTACAGGC 12420

TTTATCTTCA GCTGTTAGCG GCTTGAGTTT GCTGGATTTA CGTGGAGTTT GTACTTGAGG 12480

ATATATCTTC ATGAGCCCTT GATAACCACT GTCAGACAAG ATTTTACCAG CTTGTCCGAT 12540

ATTTCTGCGA CTCATTTTGA ACAACTTCAT ATCACGACAA TAGTTCACAG CGATATCCAA 12600

AGAAACAATT CTCCCTTGAC TTGTGACAAT CGCTTGAGCC TTCATAGCGT GAAATTTCTT 12660

TTTACCAGAA TGATTCGCTA ATTCTTTTTT TAGGGCGATT GATTTTTACT TCCGTCGCAT 12720

CAATCATTAC CGTGTCCTCA GAACTGAGAG GAGTTCTTGA AATCGTAACA CCACTTTGAA 12780

CAAGAGTTAC TTCAACCCAT TGGCTCCGAC GGATTAAGTT GCTTTCGTGA ATACCAAAAT 12840

CAGCCGCAAT TTGTTCATAA GTTCGATATT CTCGCACATA TTGAAGAGTG GCCATAAGAA 12900

GGTCTTCTAG GCTTAATTTA GGTTTTCGTC CACCTTTTGC GTGTTTAAGT TGATAAGCTG 12960

TTTTTAATAC AGCTAATATC TCTTCAAAAG TCGTGCGCTG AACACCAACA AGACGCTTAA 13020

ATCGTGCATC AGTTAGTTGT TTACTTGCTT CATCATTCAT AGAACTACTA TACCATATTT 13080

TGTTTCGCAG GAAGTCTATT GGAAAGTAAG AAATATTGAA GCTGAGGCTA TTAGAAGAAA 13140

TTGTGAGCGT GGTGCTATTT TTTCAGGTAA AATAAAATAT CACGAAGATT CACAGTTTAA 13200

AGGAGATCAC TATGTTGAAT GTTATGCTGT TTTAGATAAT ACGGTTATAG CAAGAGATAG 13260

AATAACAGTC CCTATCGATC CGTTATGTGG AAAAGATTTT ATAGAGTAGC ATATAATTGA 13320

TTCTTAACTG GAATACTCAC TATCTCTTTA CATCAAGAAA ATGACTAAAC AGGGAAGTTT 13380

GCCTTCTTCC CTTTTTTTGT TATACTAGTA GAAGAAAAAA TTAGAAAGAT TTGTGGGTGT 13440

CAAACAGCCC AGTGGGGTGT TTTAATATGG ACTTAGGTCC CACCCAAAGA GGTATTAGTG 13500

TCGTGTCTCA ATCTTATATC AATGTTATCG GTGCTGGTTT GGCAGGTTCT GAAGCAGCTT 13560

ACCAAATCGC AGAGCGTGGT ATTCCAGTTA AACTATATGA AATGCGTGGT GTCAAGTCTA 13620

CACCCCAGCA TAAAACAGAC AATTTTGCTG AGTTGGTTTG TTCCAATTCT TTGCGTGGGG 13680

ATGCTTTGAC AAATGCAGTT GGTCTTCTCA AGGAAGAAAT GCGTCGCTTG GGTTCTGTTA 13740

TCTTGGAATC TGCTGAGGCT ACACGTGTTC CTGCAGGTGG TGCCCTTGCA GTGGACCGTG 13800

ATGGTTTCTC TCAAATGGTG ACCGAAAAAG TTGCCAACCA CCCCTTGATT GAAGTGGTTC 13860

GTGATGAAAT TACAGAATTG CCGACAGATG TTATTACGGT TATCGCTACT GGTCCTTTGA 13920

GAAGTGATGC CTTGGGTGAA AAGATTGATG GTGTTTATGA TGGTCGTGGT TTTTATTTCT 13980

ACGATGCGGC AGCGCCTATT ATCGATGTCA ACACTATCGA TATGAGCAAG GTCTACCTCA 14040

AATCACGTTA TGATAAGGGA GAAGCGGCCT ACCTCAATGC CCCTATGACC AAGCAAGAAT 14100

TTATGGATTT CCATGAAGCT TTGGTCAATG CAGAAGAAGC ACCGCTTAGT TCTTTTGAAA 14160

AAGAAAAGTA CTTTGAAGGA TGTATGCCTA TCGAAGTCAT GGCCAAACGT GGCATTAAAA 14220

CTATGCTTTA TGGCCCTATG AAGCCAGTCG GTCTTGAGTA CCCAGACGAC TATACAGGAC 14280

CTCGTGATGG AGAATTTAAA ACACCTTATG CGGTTGTGCA ACTTCGTCAG GATAATGCAG 14340

CTGGTAGCCT CTACAATATT GTTGGTTTCC AGACCCACCT CAAATGGGGA GAACAAAAGC 14400

GTGTCTTCCA AATGATTCCG GGTCTTGAAA ATGCGGAGTT TGTCCGTTAT GGTGTGATGC 14460

ATCGCAATTC TTACATGGAT TCACCAAATC TTCTTGAGCA GACTTACCGT TCTAAGAAAC 14520

AACCAAATCT CTTCTTTGCT GGTCAAATGA CGGGTGTGGA AGGCTATGTT GAGTCGGCGG 14580

CTTCAGGCTT AGTTGCGGGA ATTAACGCAG CTCGTCTCTT CAAGGAAGAA AGCGAGGCTA 14640

TTTTCCCCGA GACGACAGCG ATTGGAAGCT TAGCTCATTA CATTACCCAT GCCGACAGCA 14700

AACATTTCCA ACCAATGAAT GTCAATTTTG GGATCATCAA GGAGTTGGAA GGCGAGCGTA 14760

TCCGTGATAA GAAGGCTCGT TATGAAAAAA TTGCAGAGCG TGCCCTTGCC GACTTAGAGG 14820

AATTTTTGAC TGTCTAATTT TTTTGAAAGA ATTGCTCATG ATACTATAAA AATCTTAGAA 14880

ATTGTGATAA AATAGGTAGG ATGAAAGAAG GAGAGTGAAA ATGGCGAATC CCAAGTATAA 14940

ACGTATTTTA ATCAAGTTAT CAGGTGAAGC CCTTGCCGGT GAACGTGGCG TAGGGATTGA 15000

TATCCAAACA GTTCAAACAA TCGCAAAAGA GATTCAAGAA GTTCATAGCT TAGGTATCGA 15060

AATTGCCCTT GTTATCGGTG GAGGAAATCT CTGGCGTGGA GAACCTGCAG CAGAAGCAGG 15120

TATGGACCGT GTTCAGGCAG ATTACACAGG AATGCTTGGG ACTGTTATGA ATGCTCTTGT 15180

GATGGCAGAT TCATTGCAAC AAGTTGGGGT TGATACGCGT GTACAAACAG CTATTGCCAT 15240

GCAACAAGTG GCAGAGCCTT ATGTCCGTGG ACGTGCCCTT CGTCACCTTG AAAAAGGCCG 15300

TATCGTTATC TTTGGTGCTG GAATTGGTTC ACCTTACTTC TCGACAGATA CAACAGCGGC 15360

CCTTCGTGCA GCTGAAATCG AAGCAGATGC CATCCTCATG GCTAAAAATG GTGTCGATGG 15420

TGTTTACAAT GCCGATCCTA AGAAAGATAA GACAGCTGTT AAGTTTGAAG AATTGACCCA 15480

CCGTGACGTT ATCAATAAAG GTCTTCGTAT CATGGACTCA ACAGCTTCAA CCCTCTCAAT 15540

GGACAACGAC ATTGACTTGG TTGTATTCAA CATGAACCAA CCAGGCAACA TCAAACGTGT 15600

CGTATTTGGT GAAAATATCG GAACAACAGT TTCAAATAAT ATCGAAGAAA AGGAATAAGA 15660

AAGAATATGG CTAACGCAAT TATTGAAAAA GCTAAAGAGA GAATGACCCA GTCTCACCAA 15720

TCACTTGCTC GTGAATTTGG TGGTATCCGT GCTGGTCGTG CCAATGCAAG CTTGCTTGAC 15780

CGTGTACATG TAGAATACTA TGGAGTCGAA ACTCCTCTTA ACCAAATCGC TTCAATTACG 15840

ATTCCAGAAG CGCGTGTTTT GTTGGTAACA CCATTTGACA AGTCTTCATT GAAAGACATC 15900

GAACGTGCCT TGAACGCTTC TGATATTGGT ATCACACCGG CTAATGACGG TTCTGTGATT 15960

CGCTTGGTTA TCCCAGCTCT TACAGAAGAA ACTCGTCGTG ACCTTGCTAA AGAAGTGAAG 16020

AAGGTCGGCG AAAATGCTAA AGTGGCTGTC CGCAATATCC GTCGCGATGC TATGGACGAA 16080

GCTAAGAAAC GAGAAAAAGC AAAAGAAATC ACTGAAGACG AATTGAAGAC TCTTGAAAAA 16140

GACATTCAAA AAGTAACAGA CGATGCTGTT AAACACATCG ACGACATGAC TGCTAACAAA 16200

GAGAAAGAAC TTTTGGAAGT CTAAAAATAA ACAGAAAAAC TCAGTTGGCA TTGCTGGCTG 16260

AGTTTTATTC GAAAGAAGGA AATATGAATA CAAATCTTGC AAGTTTTATC GTTGGACTGA 16320

TCATCGATGA AAACGACCGT TTTTACTTTG TGCAAAAGGA TGGTCAAACC TATGCTCTTG 16380

CTAAGGAAGA AGGCCAACAT ACAGTAGGGG ATACGGTCAA AGGTTTTGCA TACACGGATA 16440

TGAAGCAAAA ACTCCGCCTG ACAACCTTAG AAGTGACTGC CACTCAGGAC CAATTTGGTT 16500

GGGGACGTGT CACAGAGGTT CGTAAGGACT TGGGTGTCTT TGTGGATACA GGCCTTCCTG 16560

ACAAGGAAAT CGTTGTGTCA CTCGATATTC TCCCTGAGCT CAAGGAACTC TGGCCTAAGA 16620

AGGGCGACCA ACTCTACATC CGTCTTGAAG TGGATAAGAA AGACCGTATC TGGGGCCTCT 16680

TGGCTTATCA AGAAGACTTC CAACGTCTTG CTCGTCCTGC CTACAACAAC ATGCAGAACC 16740

AAAACTGGCC AGCCATTGTT TACCGTCTCA AGCTGTCAGG AACTTTTGTT TACCTACCAG 16800

AAAATAATAT GCTTGGTTTT ATTCATCCTA GCGAGCGTTA CGCAGAGCCA CGTTTGGGGC 16860

AAGTATTAGA TGCGCGCGTT ATTGGTTTCC GTGAAGTGGA CCGCACTCTG AACCTCTCCC 16920

TCAAACCACG CTCCTTTGAA ATGTTGGAAA ACGATGCTCA GATGATTTTG ACTTATTTGG 16980

AAAGCAATGG CGGTTTCATG ACCTTAAATG ACAAGTCATC TCCAGACGAC ATCAAGGCAA 17040

CCTTTGGCAT TTCTAAAGGT CAGTTCAAGA AAGCTTTAGG TGGTCTTATG AAGGCTGGTA 17100

AAATCAAGCA GGACCAGTTT GGGACAGAGT TGATTTAGGG AGGCTTATGA GAAAATCATT 17160

TTACACTTGG CTCATGACCG AGCGCAATCC TAAAAGTAAC AGTCCCAAAG CAATTTTGGC 17220

AGACCTCGCT TTTGAAGAGT CAGCCTTTCC AAAACACACA GATGATTTTG ATGAGGTCAG 17280

TCGCTTTTTG GAGGAGCATG CCAGTTTCTC TTTTAACCTA GGAGATTTTG ACAGCATTTG 17340

GCAGGAATAT CTAGAACACT AGCATTTATT CATTGGGTTT GGGCTAGTAA TTTCTCCATC 17400

CCTCTGCTAT AATAAAAAAG AATAAAAGGA TTAGAGAGGT TCTTTATTTG AAGGAACATT 17460

CAATAGACAT TCAACTGAGT CATCCAGATG ACCTGTTTCA TCTTTTTGGT TGCAATGAAC 17520

GCCATCTTCG TTTGATGGAA GAAGAGCTTG ATGTTGTGAT TCATGCTCGT ACGGAGATTG 17580

TCCAGGTTTT GGGAGAAGAG TCTGCCTGTG AGGAAGCCCG TCAAGTTATT CAGGCTTTGA 17640

TGGTCTTGGT AAATCGTGGG ATGACCGTTG GTACGCCAGA TGTAGTCACT GCGATTAGCA 17700

TGGTCAAAAA TGATGAAATT GACAAGTTTG TCGCCCTTTA CGAAGAAGAA ATTATCAAGG 17760

ATAATACTGG GAAACCTATC CGTGTCAAAA CCCTAGGGCA AAAGCTTTAT GTGGACAGTG 17820

TCAAACAGCA TGATGTGACC TTTGGAATTG GGCCAGCAGG TACAGGGAAG ACCTTCCTTG 17880

CAGTGACCTT GGCAGTGACT GCCCTTAAAC GTGGGCAAGT CAAGCGAATT ATCCTAACTC 17940

GTCCAGCGGT GGAAGCGGGA GAGAGTCTTG GATTTCTTCC GGGTGATCTT AAGGAGAAGG 18000

TGGATCCTTA CCTTCGTCCT GTTTACGATG CCTTGTATCA AATTCTTGGG AAAGACCAAA 18060

CGACTCGTCT CATGGAGCGT GAAATTATCG AAATTGCGCC CCTTGCCTAT ATGCGTGGCC 18120

GGACCTTGGA TGATGCCTTT GTCATTCTCG ATGAGGCGCA AAACACGACC ATCATGCAGA 18180

TGAAGATGTT CTTGACGCGT TTAGGTTTTC ATTCTAAGAT GATTGTCAAT GGAGATATTA 18240

GTCAGATTGA CCTGCCACGT AATGTCAAGT CCGGTTTGAT TGATGCTCAA GAGAAACTCA 18300

AGAACATCCA TCAGATTGAC TTTGTTCATT TTTCAGCCAA GGATGTGGTT CGCCATCCTG 18360

TTGTCGCTCA GATTATCCGA GCCTATGAAT ATTCTACTGA AGTTGCACAC GACTGATTTT 18420

GAGGAAGTTC GCCTGCAAAA GAATAGACTT GTTCGGTAAC TGTAAAAAGT GTTATACTAT 18480

TTTTATGGAA ACAGTATACG ACAAAGCACA AAAACTTAAC TCAAAAAACT TCAAACTATT 18540

GATTGGTGTC AAAAAGGAAA CCTTTCAACT CATGCTAGAA CACCTGAATT CAGCCTATCA 18600

GATTCAGCAC CGAAAAGGTG GACGTCCACG TAGTCTGCCC ATGGAAGACC AGCTCATTAT 18660

GACCCTCCGT TACTTGCGAT ATTATCCCAC TCAGCGTCTG CTGGCCTTTG ATTTTGGCGT 18720

CGGTGTAGCT ACGGTAAATG CCATCATCAC TTGGGTGGAG GATACACTTC GTGCGTCAGG 18780

TAGCTTTGAT TTGGACCATT TAGAAGCCCC GAGTGCTGCT GTGGCTATTG ACGTGACCGA 18840

AAGTCCGATT CAGCGTCCAA ACAAAACCAA AGCAAAAATT ATTCTGGTAA AAAGAAACGA 18900

CACACCTTAA AAACTCAAAT TATGCTGGAT TTGACGACAC ATAAAGTCTG TCAAATGGCC 18960

TTTTCTGACG GACATACGCA TGATTTTACT CTCTTCAAAG AAAGTATTGG ACAAAGTTTG 19020

CCTGAAACGA CGCTTGCCTT TGTTGACCTA GGTTATTTAG GCATCTTGAA ATTTCATGAG 19080

AATACTTTCA TTCCTGCTAA AAATTCCAAA AATCGCCGCC TGAGTGAGGA TGATAAGCAG 19140

TTAAATAAAG AGATGTCAGC GATACGAATT GAAATTGAAC ATTTTAACGC TAAATTCAAG 19200

ACCTTCCAAA TCATGTCAGT CCCTTATCGT AACCGCAGAA AACGTTTCGA GTTACGGGCG 19260

GAATTAATTT GTGCCATCAT CAATTATGAA GTGAACTAGA TTCCGAACAA GTCTAATATA 19320

CTTTTGAGAG AGGAAAATCC AGTTGTATAG GCTAAAGGTT TTATCCAAAG GTCTGAGACA 19380

ACGATTAGGC ACGATGGAAA GAACTTTTAT GTGGCTGATG ACGATCAGTG CATCTTCCTG 19440

TGTCATAATC ACAGGGCACA AGAAAGTAGG AATTTGAAAA GATGATTGAC CAACTATCTA 19500

AGTATTACAG TTGTAGGATA CTAACTGAAA AGGATATTCC AAGTATTTTA TCTTTATATG 19560

AAAGTAATCC TCTGTATTTT CAGCATTGTC CACCAGAGCC AAATTTTGCA ACTGTAAAAG 19620

AGGACATGCT TTGTCTACCT GAAGGTAAAG CTAAGGCTGA TAAGTTTTTT GTTGGATTTT 19680

GGAATGGATC TGACCTTGTG GCTGTTATGG ATTTTGTCTA TGCATATCCT GATGAGGAGA 19740

CTGTTTTTAT TGGTTTGTTT ATGGTTGATC AAGCCTATCA GAGAAAAGGG ATTGGTAGTC 19800

ATATTGTGAC AGAAGCACTA GCTTATTTTG CTAAGAACTT TCGAAAGGCA CGTTTGGCTT 19860

ATGTTAAGGG AAATCCGCAA TCTCAGCATT TTTGGGAAAA GCAGGGCTTT AAATCAATTG 19920

GATGCGAGGT TAAGCAAGAA CTCTATACGG TTGTTATCGC TGAACAGAGC CTAGAAGATT 19980

AGAAATGGCA TCAAGTAAGA ACTATTTGGA ATTTGTTTTG GAACAATTAT CAGGATTAGA 20040

TGATGTGACT TACCGTTCCA TGATGGGGGA GTATATTCTT TACTTCCGCG GCAAGATTAT 20100

TGGCGGCATT TATGACGATC GCTTTTTAGT TAAACCCGTG CAAGCAGTCT TAGATAAGAT 20160

TGACCAATCT TCTTTTGAGT TTCCATACAA AGGTGCCAAA GAAATGATTT GAGTGGAAGA 20220

ACTTGATAAT AAGATGTTTC TATAAGACCT AATTTTAGCT ATGTATAACC AACTGCCAAC 20280

GCCCAAACCT AAAAAGAAAA AGCAAGGGTG AACGAAGTAA AAAAGAAGTC TGCTAAGGCC 20340

CTGTCTTTGC ACGGGTAAAA TTTTATATAT AAAAAGAAGC TGGGACTAAA GAGCTCAGCT 20400

TCCTTTGGTT TATATAATTG TCATTACAAG ACGAAGTGGT TGGGCGAAAC TCTGTTGACT 20460

TTATTCAATT TAGAGTTTCT TATGCACAAT TGAGTCTGGA ACGAAAGTCT CCAGTTGCAA 20520

AGTATACAGT ACAATAAACC AACGATGTAA TAGCTGATGA CACAAAGCAC AGTGGGTAGG 20580

ACTTGCGAAG TCACCCTTTT CTTTTCAAAA TTTATACTAA ATCATTGATA TCAGTGTAGT 20640

CACGATTAAG TCCTTGAGCA ACTGGTAGGT TAGTCAAGTA ACCTTGATAA GTAGTCACAC 20700

CTTGACGCAA GCCTTCATCT TCAGAGATTG CTTGTGCGAA TCCTTTGCCA GCCAAAGCTT 20760

CGATATAAGG AAGAGTGACA TTGGTTAGGG CGATGGTTGA AGTGCGAGCA ACCGCACCAG 20820

GGATATTGGC AACGGCATAG TGGAGAACAC CGTGTTTTTC ATAGACGGGT TCATCGTGCG 20880

TTGTCACACG GTCAGCTGTT TCGATAACGC CACCTTGGTC AACAGCAACG TCAACGATAC 20940

AGAGCCTGGA CGCATTTGTT TGACCATCTC ATCTGTCACC AATTCCGGTG CTTTTGCACC 21000

AGGGATGAGA ATGGCTCCAA TCACCACATC AGCATCTCTC ACACTTGCTT CAATGTTGAA 21060

TGAATTAGAC ATAAGAGTTT GAATTTGACT TCCAAAGACT TCTTCTAGAA CTGAGAGACG 21120

CTTGGAACTA ATATCTAAAA TAGTCACTTG AGCACCAAGA CCAAGGGCGA TGCGGGCAGC 21180

ATGTGTACCG ACGACACCAC CACCGATGAT AGTTACTTTT CCTTTTGGAA CACCTGGTAC 21240

ACCACCAAGT AGAACACCAG AGCCACCAGC TTGCTTAGTA AGGAAGTGAG CTCCGATTTG 21300

AACAGCCATA CGACCTGCAA CCTCACTCAT AGGAACGAGG AGCGGTAGTT GTCCTTGATT 21360

GTCACGAACA GTTTCAGTTG TTTTTGCTGT TAACATAGCA TCTGCTAATT CTGGAGCAGC 21420

GGCCATGTGC AAGTAGGTGA AGAGAAGAAG ATCGTCGCGC AAGTAACCGT ATTCAGAACT 21480

TAAAGATTCT TTTACTTTCA CAACCAACTC TGCTGCCCAA GCTTCACCAG CAGTAGCGAC 21540

AATCTCAGCT CCTTGCTTTT GATAGTCAGC ATCAGTAAAG CCAGAACCGA GACCAGCATT 21600

TGTTTCGATA AGGACACGAT GACCACGACT AACTAAGCTA TGAACACCTG CAGGTGTGAG 21560

GGCGACACGG TTTTCGTTAT TTTTAATTTC TTTTGGGATT CCGATTAACA TTGAGATAAC 21720

CTACCTTTCA ATTGACGGTC TTGTTTTGGT TGTCACATTC CAGTTCATAA ATCAAAAATG 21780

TGACGGTTTC ATTGTATATG AAACCGCTTC AAAAATCAAG AAAAACTTGT CATCCAAATT 21840

TTTTTATGCT AGACTAGTGA AAATCAAGCT CTAATGGAGG GAAAAGTATG GAATCAATAT 21900

TTGTGAAATT TGCCCAGTAT CCGTCTATAG AAACGGAGCG TTTATTGCTC AGACCTGTAA 21960

CTTTGGATGA TGCGGAAcAA TGTTTGACTA TGCCTCGGAC AAGGGTAATA CACGTTACAC 22020

TTTTCCAACC AATCAAAGCT TGGAAGAAAC CAAGAATAAC ATTGCTCAGT TCTACTTGGC 22080

TAATCCCTTG GGACGTTGGG GAATAGAACT AAAAAGCAAT GGTCAGTTTA TTGGAACCAT 22140

TGACTTGCAC AAGATTGATT CTGTTCTTAA GAAGGCAGCT ATTGGCTACA TTATCAATAA 22200

AAAGTATTGG AATCAAGGAT TAACGACAGA AGCCAATCGT GCTGTGATTG AGCTAGCTTT 22260

TGAGAAGATA GGGATGAATA AGTTGACTGC CCTTCACGAT AAGGCTAATC CCGCGTCAGG 22320

AAAGGTCATG GAGAAATCAG GCATGCGTTT TTCCCATGCA GAACCATATG CTTGTATGGA 22380

CCAGCATGAA AAAGGCCGAA TCGTGACAAG AGTTCATTAT GTCTTGACCA AGGAAGACTA 22440

TTTTGCAAAT AAATAAGCAG TTGAAAAGAA ATTTTTCGAC TGTTTTTTCT TCCTCTTACG 22500

AATAATCTAA GAGAGGAGAA AATATGGAAG CAATTATCGA GAAAATCAAA GAGTATAAAA 22560

TCATCGTCAT CTGTACTGGT CTGGGCTTGC TTGTAGGAGG ATTTTTCCTG CTAAAACCAG 22620

CTCCACAAAC ACCTGTCAAA GAGACGAATT TGCAGGCTGA AGTTGCAGCT GTTTCCAAGG 22680

ACTCATCGAC CGAAAAGGAA GTGAAGAAGG AAGAAAAGGA AGAACCCCTT GAACAAGATC 22740

TAATCACAGT AGATGTCAAA GGTGCTGTCA AATCGCCAGG GATTTATGAC TTGCCTGTAG 22800

GTAGTCGAGT CAATGATGCT GTTCAGAAGG CTGGTGGCTT GACAGAGCAA GCAGACAGCA 22860

AGTCGCTCAA TCTAGCTCAG AAAGTTAGTG ATGAGGCTCT GGTTTACGTT CCTACTAAGG 22920

GAGAAGAAGC AGTTAGTCAA CAGACTGGTT CGGGGACAGC TTCTTCAACA AGCAAGGAAA 22980

AGAAGGTCAA TCTCAACAAG GCCAGTCTGG AAGAACTCAA GCAGGTCAAG GGACTGGGAG 23040

GAAAACGAGC TCAGGACATT ATTGACCATC GTGAGGCAAA TGGCAAGTTC AAGTCAGTAG 23100

ACGAGCTCAA GAAGGTCTCT GGCATTGGTG GCAAAACAAT AGAAAAGCTT AAAGACTATG 23160

TTACAGTGGA TTAAGAATTT CTCTATTCCC CTAATTTACC TGAGTTTTCT ATTACTTTGG 23220

CTTTATTACG CTATTTTCTC AGCATCTTAT CTTGCTTTGT TGGGCTTTGT TTTTCTGCTA 23280

GTCTGTCTCT TTATCCAATT TCCGTGGAAA TCTGCTGGTA AAGTTCTAAT AATTTGCGGA 23340

ATCTTTGGAT TTTGGTTTGT TTTTCAAAAT TGGCAACAGA GTCAAGCGAG TCAAAATCTG 23400

GCGGATTCTG TTGAAAGGGT ACGGATTTTG CCTGATACTA TTAAGGTTAA TGGTGATAGT 23460

CTATCCTTTC GTGGCAAGTC TAACGGTCGT GCTTTCCAAG TCTATTATAA ACTCCAGTCC 23520

GAGGAGGAGA AAGAAGCCTT TCAAGCTTTA ACTGACCTGC ATGAGATAGG ACTAGAAGGG 23580

AAGCTTTCGG AGCCAGAAGG GCAGAGAAAT TTTGGTGGCT TTAATTACCA AGCCTATCTG 23640

AAGACTCAGG GAATTTACCA GACTCTCAAT ATCAAAACAA TCCAGTCACT TCAAAAGATT 23700

GGCAGTTGGG ATATAGGAGA AAACTTGTCC AGTTTACGTC GAAAGGCTGT GGTTTGGATT 23760

AAGACGCACT TTCCAGACCC TATGGGCAAT TACATGACAG GACTCTTGCT GGGACATCTG 23820

GACACCGACT TTGAGGAGAT GAATGAGCTT TATTCCAGTC TAGGAATTAT CCACCTCTTT 23880

GCCCTATCTG GCATGCAGGT AGGTTTTTTC ATGAATGGAT TTAAGAAACT TCTCTTGCGA 23940

TTGGGCTTGA CCCAAGAAAA GTTGAAATGG CTGACTTATC CCTTTTCCCT TATCTATGCG 24000

GGACTAACTG GATTTTCAGC ATCGGTTATT CGCAGTCTCT TGCAAAAGCT ACTGGCTCAA 24060

CATGGGGTTA AGGGCTTGGA TAATTTTGCC TTGACGGTGC TTGTCCTCTT TATTGTCATG 24120

CCAAACTTTT TCTTGACAGC AGGAGGAGTC TTGTCCTGCG CTTATGCTTT TATCCTGACC 24180

ATGACCAGCA AAGAAGGGGA GGGGCTCAAG GCTGTTACTA GTGAAAGTCT AGTCATCTCC 24240

TTGGGCATAT TGCCCATTCT ATCCTTCTAT TTTGCGGAAT TTCAACCTTG GTCTATCCTT 24300

TTGACCTTTG TCTTTTCCTT TCTTTTTGAC TTGGTCTTCT TACCGCTCTT GTCTATCTTA 24360

TTTGTCCTTT CCTTTCTCTA TCCAGTCATT CAGCTGAACT TTATCTTTGA ATGGTTAGAG 24420

GGCATTATTC GCTTGGTCTC GCAGGTGGCA AGGAGACCAC TTGTCTTTGG TCAACCCAAC 24480

GCATGGCTTT TAATCTTATT GTTAATTTCC TTGGCTTTGG TCTATGATTT GAGGAAAAAC 24540

ATTAAAGGAT TAACAGTATT GAGTTTATTG ATTACAGGTC TCTTTTTCCT TACCAAGTAT 24600

CCACTGGAAA ATGAAATCAC CATGCTGGAT GTGGGGCAAG GAGAAAGTAT TTTCTACGGG 24660

ATGTAACTGG GAAAACCATT CTCATAGATG TAGGTGGTAA GGCAGAATCT TATAAGAAAA 24720

TCAAAAAATG GCAAGAAAAG ATGACGACCA GCAATGCCCA GCGAACCTTG ATTCCCTATC 24780

TCAAAAGTCG AGGAGTAGCT AAGATTGACC AGCTAATTTT GACTAACACG GACAAGGAGC 24840

ATGTTGGAGA TTTGTCAGAG ATGACCAAGG CTTTCCATGT AGGGGAGATT CTAGTATCAA 24900

AAGACAGTCT GAAACAGAAG GAATTTGTGG CAGAACTACA GGCGACTCAA ACAAAGGTGC 24960

GTAGTATGAT AGTAGGGGAG AACTTGCCCA TTTTTGGAAG TCAGTTAGAA GTTCTATCTC 25020

CAAGGAAAAT GGGAGATGGA GGACACGATG ATACCCTAGT TCTGTATGGG AAATTCTTGG 25080

ATAAGCAATT TCTCTTCACG GGAAATTTGG AGGAGAAAGG AGAGAAGGAC TTGCTGAAGC 25140

ACTATCCAGA CTTGAAAGTA AATGTTTTGA AAGCTAGCCA ACATGGCAAT AAAAAATCAT 25200

CAAGTCCAGC CTTTCTAGAA AAACTCAAAC CAGAGCTTAC TCTTATCTCA GTTGGAAAGA 25260

GCAATCGAAT GAAACTCCCC CATCAGGAAA CATTGACACG ACTGGAAGGT ATCAATAGCA 25320

AAGTTTATCG AACTGACCAG CAAGGAGCTA TACGTTTTAA GGGGTTGGAT AGTTGGAAAA 25380

TCGAAAGTGT TCGATAGGAA GGATAAATGT TGTAGATTAG TGAAATAAAC TAAAAATTTG 25440

TTGCATAATA ATGATAAAAA TGGTATAATG AAAACGTATT CAATATTGAG GATATAAAAT 25500

CATTAAAAAT CAGCAAAAGT TGTTTTATTA GTTAGTTTAT AATCTATTGG TCTTCTTCAG 25560

TCCAGTGTAT CTGCTGTGAC AGTCACTAAA AGTTACAAGT ATGATTGGAA TACGGTTTGG 25620

GAATATAGTA CCAACTATCA CGACCATCAG TATGCTTGGA TTCCGTCATG GTCTCGTTAT 25680

GACAGCTATT CTGAGTATAA AGTTGGCGGA GGCTGGAACT ACGCTCGTTA TGAGGTCATA 25740

AACTATTACA GCGGAGGCTA TTAATTCTTA AAGAGTGAGA AAAAGGAGGG CTAGATATGT 25800

TGCAGCTTAC TCATGTGACC TTAAAAACGC GACAAGTCAT CTTGCAAGAT GTGGATTTCA 25860

CCTTTAAAAA GGGTAGGGTT TATGGTCTTC TTGCTATCAA TGGCTCTGGA AAGACGACCC 25920

TGTTCCGTGC CATTAGCAAT TTAATTCCCA TAAGTAGTGG AAATATCGCA GCCCCTCCTT 25980

CTTTATTTTA TTATGAGAGT ATTGAATGGC TGGATGGAAA CTTAAGTGGG ATGGACTACC 26040

TTCGTCTTAT CAAAAACATC TGGAAGTCAG GTCTGAACTT GAGGGATGAA ATCGCCTATT 26100

GGGAAATGTC TGACTATATC AGTCTTCCCA TTCGCAAGTA TTCCTTAGGC ATGAAGCAAC 26160

GCTTGGTGAT TGCCATGTAT TTCCTCAGTC AGGCCAAATG CTGGCTCATG GATGAGATTA 26220

CAAATGGCTT AGATGAGTAT TATCGACAGA AGTTTTTTGA TAGGCTAGCA CAAATCGATA 26280

GACAAGAACA GCTGGTTCTT TTAAGTTCCC ACTATAAGGA AGAGTTGGTT GATGTCTGCG 26340

ATAGAGTAGT AACCATTCAT CAGGGGCAGA TAGAAGAGGT TTAGTTTATG AAAGATGTTA 26400

GTCTATTTTT ATTGAAAAAA GTTTTCAAAA GCCGCTTAAA CTGGATTGTC TTAGCTTTAT 26460

TTGTATCTGT ACTCGGTGTT ACCTTTTATT TAAATAGTCA GACTGCAAAC TCACACAGCT 26520

TGGAGAGCAG GTTGGAAAGT CGCATTGCAG CCAACGAGAG GGCTATCAAT GAAAATGAAG 26580

AGAAACTCTC CCAAATGTCT GATACCAGCT CGGAGGAATA CCAGTTTGCT AAAAATAATT 26640

TAGACGTGCA AAAAAATCTT TTGACGCGAA AGACAGAAAT TCTGACTTTA TTAAAAGAAG 26700

GGCGCTGGAA AGAAGCCTAC TATTTGCAGT GGCAAGATGA AGAGAAGAAT TATGAATTTG 26760

TATCAAATGA CCCGACTGCT AGCCCTGGCT TAAAAATGGG GGTTGACCGC GAACGGAAGA 26820

TTTACCAAGC CCTGTATCCC TTGAACATAA AAGCACATAC TTTGGAGTTT CCGACCCACG 26880

GGATTGATCA GATTGTCTGG ATTTTAGAGG TTATCATCCC AAGTTTGTTT GTGGTTGCTA 26940

TTATTTTTAT GCTAACACAA CTATTTGCAG AAAGATATCA AAATCATCTG GACACAGCTC 27000

ACTTATATCC TGTTTCAAAA GTGACATTTG CAATATCCTC TCTTGGAGTT GGAGTGGGAT 27060

ATGTAACTGT GCTGTTTATC GGAATCTGTG GCTTTTCTTT TCTAGTGGGA AGTCTGATAA 27120

GTGGTTTTGG ACAGTTAGAT TATCCCTACC CAATTTATAG CTTAGTGAAT CAAGAAGTAA 27180

CTATTGGGAA AATACAAGAT GTATTATTTC CTGGCTTGCT CTTAGCTTTC TTAGCCTTTA 27240

TCGTCATTGT GGAAGTTGTG TACTTGATTG CTTACTTTTT CAAGCAAAAA ATGCCTGTCC 27300

TCTTTCTTTC ACTCATTGGG ATTGTTGGCT TATTGTTTGG TATCCAAACC ATTCAGCCTC 27360

TTCAAAGGAT TGCACATCTG ATTCCCTTTA CTTACTTGCG TTCAGTGGAG ATTTTATCTG 27420

GAAGATTACC TAAGCAGATT GATAATGTCG ATCTAAATTG GAGCATGGGA ATGGTCTTAC 27480

TTCCTTGCCT GATTATCTTT TTGCTATTGG GAATTCTATT TATTGAAAGA TGGGGAAGTT 27540

CACAGAAAAA AGAATTTTTT AATAGATTCT AGCTTTCCTA TAGGTAGGGA AAATAAGTAA 27600

AAACTAACAT AGAGAGGGAA TCAACTTGAT TCTCTCTTTT TGATTCGAAA ACCAAACCAA 27660

AATACAAACA CAAACTTTTC AAAAAATAAC TTTTTATCTT GACAAGAGCT AGAAAACTTG 27720

GTATCATATA AAAGTTGAGA AAAGCAGAAG TGAGAGCTTC TCGCCTTGTG ACATTAAGTT 27780

GCCTGGCCCT ACGGATGAAA AGTTTCGAAG AAACGCTATC ATAACGTGCG GGCTTGTATA 27840

TTTACAAGTC CGCTATTGTT TTTCTCTAAT AAAACAAAAG AGGTGAAAAC CATAGCAAAG 27900

CAAGACTTAT TCATCAATGA TGAGATTCGT GTACGTGAAG TTCGCTTGAT TGGTCTTGAA 27960

GGAGAACAGC TAGGTATCAA GCCACTCAGT GAAGCGCAAG CTTTGGCTGA TAACGCTAAT 28020

GTTGACCTAG TATTGATTCA ACCCCAAGCC AAACCGCCTG TTGCAAAAAT TATGGACTAC 28080

GGTAAGTTCA AATTTGAGTA CGAGAAGAAG CAAAAAGAAC AACGTAAAAA ACAAAAGCGTT 28140

GTTACTGTGA AAGAAGTTCG TCTAAGTCCG G 28171

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7147 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CCGCTCAACT TTTGCAATCA AGGCTAAGTA GACAGCAGCA AATTTCATAT TGTATAATTT 60

CTGACTCATA CTTCTCTCTT TCTATGTGTA CTAGTATAAA TAAGAAAAAG AAGGCCGTCA 120

AGCCTTCTTT TGATTTATTC TTCTGCTTCA TCTTCTGTAA ATTGACTATT GTACAAGTCA 180

GCGTAGAAGC CACCTTGCGC CATCAGTTCC TCATAGTTGC CTTGCTCGAT GATATTTCCA 240

TCTTTCATGA CCAAGATCAA GTCTGCATTT CGGATGGTTG ACAAGCGGTG GGCAATGACA 300

AAGGATGTGC GTCCTTCCAT CAAACGGTCC ATGGCTTTTT GGATCAATTC CTCTGTCCGT 360

GTGTCAACAG AAGAAGTCGC CTCATCCAAA ATCAAAAGCG GTGCATCCTT AAGAAGGGCA 420

CGAGCAATAG TCAATAGTTG TTTTTGTCTT ACAGACAAGG TCACGGTGTC ATCCAAGATG 480

GTATCATAGC CATCTGGCAA GGTCATAATA AAGTGGTGAA TTCCCACAGC CTTACTAGCT 540

TCCATCATTC GTTCATCACT AATCCCTATT TGATTATAGA TGAGATTGTC TCGAATAGTT 600

CCTTCAAAGA GCCAGGTATC CTGCAAGACC ATTGAAAAGG CATCATGCAC TTCTGAACGC 660

GTCATAGCCT TGGTATCCAC ACCATCAATG CGAATACTTC CCTTATCAAT CTCATAGAAT 720

TTCATCAAAA GATTGACAAT GGTTGTCTTA CCAGCCCCAG TCGGCCCAAC AATGGCAACC 780

TTTTGACCAG CATGAGCTGT CGCAGAGAAG TCATAGTCTT GAACATTGAC ACCGTCCACC 840

AGAATTTCTC CTGCTGACAC GTCGTAGAAA CGTGGAATCA GATTGACCAG AGTTGATTTA 900

CCAGAACCTG TTGACCCAAT AAAGGCCACT GTTTGACCAG TTTCTGCTTT AAAGCTAACA 960

TGTTCAATAA CTGCCTCCGA ATTTGCCGCA TAGCGgAAGG TCACATCCTT AAACTCGACC 1020

TGACCTTTGA AGTTTTCATC AGTCAGCTGC ACTTGAACAG GGTTTTGGAT AGAAGAATGC 1080

AAATCTAAAA CTTGATTAAT CCGCTTAGCA GAGACCATAG TTCGGGGAAG AACGATGAAG 1140

AGTGCTCCCA TGAGAAGGAA GCCCATGACA ACCTACATGG CATAAGACAT GAAAACAATC 1200

ATGTCACTAA AGAGAGGCAG ACGCGCTATC GGAGCAGCGT CGTTAATCAC ATAGGCCCCA 1260

ATCCAGTAAA TCGCCACACT CAAACCACTT GAAATCCCCA TCATGATAGG ATTCAAAATA 1320

GCCATAAGAC GGTTGACAAA CAAATTCAAA CGGGTCAATT CATCATTTAC TGCTGCAAAT 1380

TTTTCATTTT GATAATCCTC TGCATTGTAG GCACGAACGA CACGAATACC TGTTAAACTC 1440

TCACGAGTGA TACTGTTCAG TTTATCTGTC AGCCCCTGAA TCAAGGACTG TTTTGGAAAG 1500

GCTAGCGTCA TCAAAACGGT CGTCATCAGG ACGTTGATAA TCACTGCCAC AAGTACGGCC 1560

CAGAGCCAGT ATTCTGAATG ACCTAAAATC TTCCCAATAG CCCAGATAGC CATAATTGAA 1620

CCACGCGTTA CCACTTGCAA GCCCATAGTA ATCAACATTT GAACTTGAGT AATGTCATTG 1680

GTAGTACGCG TCAAGAGGCT AGGAATTGAA AATTTCTTAA TCTCTGTCTG CGAGTAATCC 1740

AAAACTCGGT TAAAAATATC ACTTCTCAGC CTACTAGTAT AAGAAGCCGC CACTCGGGAT 1800

GCAAAAAATC CAACTGCAAC TACGGACAAG AAGGCAAGAA AGGACATTCC CATCATCATG 1860

CTTGCCGACT GCCACAACTC ATCTAAATTA GTTTCTTGAC TACCTAGCAA ATCCGTAATT 1920

TTCGAGATAT AGGTCGGCAC TTCCAACTCT AGATAGACCG AAAAGCAAGT AAAGAGAATG 1980

GCTAGTAAAA TCATCCCCCA TTCTTTTCTA CTAATTCTTT TGGCTAATTT CTTTATTCTC 2040

TCCTCCTATT CCCTTGATAT TTTGCCTGTA GTTGACCGAG AACCTTCTCA AAAATCAGTA 2100

ATTCATCTTC ATCAATGTCT TCCATCAACT GCTTGTCTAT GCGTTCAAAA AAAGCCTTAA 2160

CCTGTTGCAT CTGAGAACGT GCTTTGTCCG TCAGACGAAC AAACTTAGCC CGCTTATCAA 2220

CAGGACTCGC CTCCAATTCC ACCAAACCAT TTTGCACTAT ACGCTTAACC AGATTACTAG 2280

CAACAGGCTT GGTAATATTG AGTTCCTGCT CGATATCTTT AATCAAGACC AAGTCTTGGT 2340

TTTTCTCGCG ATTATCCAAA AAACGCACAA CCTGACCTTG CGGCCCACCC ATAAATTCAA 2400

TGCCGCAACG TTTGGCTTCC TTTTGCACCA TCAGGTGAAT TTGATGACCA AAACGCTTAA 2460

AGACTAACAT CGGTTTATCC ATAATCTCCC CCTTCTAAAT AAAAATAGTT CTCTGGAGAA 2520

TAATTAAATT TCTATGAGAA CTATTTTCTT GATTAAAAAA ATCCCAAGTG ATTTTCTCAC 2580

TTAGGATCAT GTTCTATAGG TTAAATTAAA ACCCATCTAC GTTCGTATAA ATCTTTTGGA 2640

CGTCTTCGTC GTCTTCAAGA ACGCTGTAAA GTTTTTCAAA GGTTTCAAGG TCTTCGCCTG 2700

ACAATTCCAC TTCTGACTGA GGAATCATTT CCAATTCAGT CACTTGGAAT TCTTCAATAC 2760

CAGACTCACG GAGGGCAACG ATAGCCTTGT GAAGGTCAGT TGGCGCTGTG TAAACTGTGA 2820

TTGTACCTTC TTGTGCTTCT ACGTCATCCA CATCCACATC CGCTTCGAGC AATTGCTCAA 2880

AGACTGCGTC CGCATCTTCA CCTCCAAATA CAATAACACC TTTGTTGTCA AAGAGGTAAG 2940

AAACAGAACC TGAAGCGCCC ATGTTTCCGC CGTTTTTACC AAAGGCTGCA CGGACATTGG 3000

CTGCTGTACG GTTGACGTTA GAAGTCAAAG TATCCACAAT TAGCATAGAG CCATTTGGCC 3060

CAAAACCTTC GTAACGTCCT TCTGTAAAGG TTTCGTCTGT GTTTCCTTTG GCTTTATCAA 3120

TCGCTTTATC GATAATGTGT TTTGGCACTT GGGCTTGTTT AGCACGGTCG ATAACGAATT 3180

TCAAAGCTGA GTTTGATTCT GGATCTGGAT CACCTTTTTT AGCTGCTACA TAGATTTCTA 3240

CACCAAATTT TGCATATACT TTAGAGTTAG CTCCATCTTT AGCCGTTTTC TTGGCTACGA 3300

TATTGGCCCA TTTACGTCCC ATTAGGAATC TCCTTTTTTC ACATTTTAAT CTTTCTTATT 3360

ATAACACAAG TTTTTTTGAT TTTCACTAGA GGAAATGGAT TTTATTAGCA AATCAAGCTA 3420

GGATAGCACT TTACCTGCTA AGATGGTCTT GCCTTTCTAT CTTTATCAAC AGGCACTCAT 3480

CCACATTCAA AAAACAAACT AGACCATTAT CTGCAAATAG AAAGTTTCAG CCAAGTTTGA 3540

CAAAGTCAGC TCAAATTACT GTTTGAAGTT TGTAGATATA AGCGACAAAA ACAATCATAC 3600

TGCACCTTTT GTTGACAGTC TACTCCAGAC ATATCATAGT TCAAGTAAAT ACTTTGAAAT 3660

TCAACAGTTC TTATAGGCGC TATTGTATTC TAAGAAATCA ATAGAACAGT TTCTAAGCAA 3720

ACCTCTAATA CTCAATAAAA ATCAAAGAGC AAACTAGAAA GCTAGCCTCA GGTTGCTCAA 3780

AACACTGTTT TGAGGTTGCG GATGGGGCTG ACATGGTTTG AAGAGATTTT CGAAGAGTAT 3840

AATTTACGTG TTCCCAAGAT GGAGAAGTTA GACTAGTACA CTGGCACTTC TAAAACATTG 3900

CTAGCAATTG ATTTGTTCAT ATTTAATTTC ATTTTTTCCA TAAATGGGTA TTAGATATAA 3960

ACAGCAAAAT ATTTCCGATA CGTGTCGTTC TTGAATTTCC AATCATCTAA AACAAGTAAA 4020

GGATAATCAA TCCCCTGTAT ATCAAGGAAT TGGCTACCCT TTTTACTTTT TTACACATTC 4080

TGTTTGATAG ATTCATTTTA ACATCACGAG CATACTCCAA TGGAAATCGC TAGGCAAGAG 4140

ATAAACTTTC AGATATCCGC AGAGAGATCA TCGCCTCTTT TTGTCGCAAG CATTCTCCTC 4200

TCCTAGTCAT TTTCTACCTT ATCTTCTACC TGAGGATAGA GAGTTGTTCC CCAAATAGAA 4260

ATCGTCCGCT TACGCACTAG TGGCAAATCG GTTTTTTCAT AAACCGTACG CCACCATTCC 4320

CAGGCAAGCC CGGTACACTC TCTAATTTTG ACAGAGAGAT TACGAACATT CCCTTTTAAA 4380

GGAATACTAG TGGTAAAGTG AGCCGTTAAA TCCTGCCCAT TTCTGTCCCA AGCCTTAGGA 4440

GTCAAGACTT CCTTACCTTG ATGATCATAG GATAATTCAT TCCAAGTAAT ATAATATTGG 4500

GCAACATAGG CACCACTATG ATCCAGCAGT AAATCTCCGT TTCTGTAAGC TGTAACCTTA 4560

GTCTCAACAT AGTCTGTACT ATTTTGAAAG GTCGCAACTA CATTGTCACG TAAAAAAGAA 4620

GTTGTATAGG AAATCGGCAA GCCTGGATGA TCTGCTGTAA AGCGACTGCC TTCTTGAATC 4680

AAGTCCTCTA CCATATCCAC CTTGCCTGTT ACAACTCGGG CACCCGAACT TGGGTCGCCC 4740

CCTAAAATAA CCGCCTTCAC TTCTGTATTG TCCAAAATCT GTTTCCACTC TGTCTGAGGA 4800

GCTACCTTGA CTCCTTTTAT CAAAGCTTCA AAAGCAGCCT CTACTTCATC ACTCTTACTC 4860

GTGGTTTCCA ACTTGAGATA GACTTGGCGC CCATAAGCAA CACTCGAAAT ATAGACCAAA 4920

GGACGCTCTG CAGAAATTCC TCTCTGTTTT AAATCCTCTA CCGTTACAGT ATCTTGAAAC 4980

ACATCTCCTG GATTTTTAAC AGCATCTACG CTGACTGTAT AATAAATCTG CTTAAAATTA 5040

ACAATCTGAA TCTGCTTTTC GCCTGAATGG ACAGAGTTAA AATCAATATC AAGAGAATTC 5100

CCTGTCTTTT CAAAGTCAGA ACCAAACTTG ACCTTGAGTT GTTCCATGCT GTGAGCCGTG 5160

ATTTTTTCAT ACTGCATTCT AGCTGGGACA TTATTGACCT GACCATAATC TTGATGCCAC 5220

TTAGCCAACA AATCGTTTAC CGCTCCGCGA ACACTTGAAT TGCTGGGGTC TTCCACTTGG 5280

AGAAAGCTAT CGCTACTTGC CAAACCAGGC AAATCAATAC TATAAGTCAT CGGAGCACGA 5340

TCGACCGCAA GAAGAGTGGG ATTATTCTCT AACAAGGTCT CATCCACTAC GAGAAGTGCT 5400

CCAGGATAGA GGCGACTGTC GTTGGTAGCT GTTACAGAAA TATCACTTGT ATTTGTCGAC 5460

AAGCTCCGCT TCTTTCTTTC GATAACAACA AACTCATCGG GTAGCTGATT ACCCTCTTTG 5520

ATGAAACGAT TTTCAATACT TTCTCCCTGA TGGGTCAAGA GTTTCTTTTT ATCGTAATTC 5580

ATAGCTAGTA TAAAGTCATT TACTGCTTTA TTTGCCATCT TCTACCTCCT AATAAGTTCC 5640

TGGATTGAGT TGCATAAACT CAGACTTGTT CAGCGAAATC AGCCGTGGTT GGACTAAGTA 5700

ATCCAAAATT TCCTCGTACA ATTCTTCTGA GACATTGCGT CGCCGTCTGG CTAAATAAGA 5760

AGTCGGAATG ACCGTATTAT CCAACATAAA TACCTTATCT AAGTCAATCA AGGTTGGTCT 5820

TGTAAAAGGA TTACGAGCTA GATCCGGCTC TTCTATCATA AAGTTCTTGA CCAAACGTCT 5880

GGTCAAGAGA GCTGGTTTGA AGGTCTGATT TTTAACCAAC TCTTTGTTTT TAGTCATGCT 5940

GTTGTCAATA CAGATATACA TATGATTCTT CACAGCCAAA TCGCTACTAA TAGTCGGAAA 6000

AGGCAAATAA AGAGCTACAA CATCTCCTCT CTTAATCAAG CAAGAGCACC CCCTTTTCTC 6060

CTAATGTAAC ATAGACAGGA TTGACCAAGT CTTCTGATTG ACTCAGAATT TCCAAAGTTT 6120

GAGTTTGGCG CGCTGTCAAT TTAGTAGCAT CTTGTCTCTT CAATACAAAA TGCTTGTCGC 6180

CAATAACCTT GACAATATAA TCCTTCTCCA AAGCTGACTG GTAAATCCAC ATCAGATGTT 6240

GTCTGTCCTG AGAACTCAAG AGAGAAGGAT TTTCAAGCCT CCCGATAGTC TGATAAAAAT 6300

CAAAAACAGG AGCTAACTCC TGCCAATCTG ATTGGCTAGT TGTCAAGGCT AGAAAAAGGG 6360

CTTTGCGAGC TGATACTTCT TGGTTAGCCT TGAGAGTTAC TTTCCCCTCC AAGTTTTTTA 6420

GAAATCGGGA AACTCCAGAA AGCAAATTTT TCTCTAACTG CGAGAAATAA AAACCTTTCG 6480

TTCCCAGACA TAAGTCTTTC ATGTCGCTTT CTCTAGCAAA TAAGAGCTCA AACATTTGAT 6540

AGTAAAAGAA AAATATCTGG CACTGGGTCG CGCTCATCTT TTCCTTATCG GCTTCTTTTT 6600

TTAACCAGAG CAAGGGCGAC AGGTAGCTGG ATTGAGACAT TTCCTCTACC TCCTACTCTT 6660

TTTTAACTGG AGCATCTGCA CTAGCTGCCA CTTCTTTTGA CTGGATACTT TCCCACTGGT 6720

TAATCTCCTC TGAGATAAGA CCTTCGCATG TCTTGACAAA TAGGGCAAAA GCCTTGGTCT 6780

TTCCTGCATA TTTCTCCGTT TGGCATTGAT AGAGGAATTT TTCTTTCTCC AGGAGTTGCG 6840

CAGTTTTTTG GTAAGAAATC CAATTTTCCT TTGCATTATA CAAATTGATA ATCCCCTCAC 6900

ACAGCAAGCC GAGACTGGAT AAGGCAACCG AAATCAAACG GTAGCGATCA CCTGGCATAG 6960

GAATAGCACA AAAGACAGCT ATGAGGAAAC CTGCCACGAT TTCTGTTATT TTTAATACCT 7020

TATAGCGCCT ACGATGTTGA ACGCTTTTCT TTAAAAAATG AGCTATCTGT ACGTCTAATC 7080

GCTCTGTCAG GTACATTTCT TCTGGCGTCA TATTCGTAAC TCCTTTCATT TACTTTGATA 7140

ATCAGGG 7147

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 755 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

CCGCATGGGA TTGGTGTCCT TTTGGGCAAT CTCTTTGACC AAACTGGAAA CATGTTTTAT 60

GCGCCTGCCT TTACTGCCCT TGTCGGCGGT ACGTCTATAT GATCCTAGTC GCAAAAGTTC 120

CCCGCTTTGG AGCCATTACC ACTATCGGCC TTGTCATTGC CCTCTTTTTC TTGGGAACTA 180

AACACGGTGC TGGTTCCTTC CTTCCTGGAA TTATCTGTGG CCTCCTAGCA GATGGAGTAG 240

CTCATTTAGG AAAATACAAG GACAAAACAA AGAACTTCCT TTCTTTCATT ATTTTCGCCT 300

TTAGTACAAC AGGACCAATC TTGCTTATGT GGATTGCGCC CAAAGCCTAT ATGGCTACTC 360

TTCTGGCAAG AGGAAAATCC CAAGAATATA TCGACCGTAT CATGGTCGCT CCAAACCCTG 420

GAACTGTCCT TCTATTTATC GCAAGTATTG TCATCGGAGC CCTAGTGGGT GCCTTGATTG 480

GACAAGCCTT GAGTAAAAAA TTTGCCCAGA AAATCTGATC AGTTAAAAAG AGCCACGCGG 540

CTCTTTTTTA TTTATGGCTC AATTTCTTAG TCAAGAAATC TCCCAAGAAT TGGATTGCAA 600

AGATAATCAA AATGATAATA ATGGTTGCCA AGATGGTCAC ATCGTGATTG TAGCGGTTAA 660

ATCCATAAGC GATGGCTACG TTACCGATAC CACCAGCTCC AACCGCACCG GCCATAGCTG 720

TTtcCCAACA AGGGaAtCAA GGTcACAGTC GTCAC 755

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3010 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

TTCAATTGGT ATCTCAATCA ACGGTCTTCA CATGGTTTCA ACTGGTTTGA CTCTTGAAAA 60

AGCGAAAGCT GCTGGTTACA ACGCAACTGA AACAGGCTTT AACGATCTTC AAAAACCAGA 120

ATTCATGAAA CATGACAACC ATGAAGTAGC AATTAAGATT GTCTTTGACA AAGATAGCCG 180

TGAAATTCTT GGTGCCCAAA TGGTTTCACA TGATATTGCA ATTAGCATGG GAATCCACAT 240

GTTCTCACTT GCTATCCAAG AGCATGTGAC AATTGATAAA TTGGCATTGA CAGACCTCTT 300

CTTCTTGCCA CACTTCAACA AACCATACAA CTACATCACA ATGGCTGCCC TTACGGCTGA 360

AAATTAAAAA TGAATGAGCT ATCTGGCCTT AAGTTAAGGT CAGATAGTTT TTAGCTAATT 420

TGTCCCCATA CAATTATAGT TTTTTTATCT TGTGCTTCAT TCTGTTCTGA CTTAAAATGA 480

AAAGGTAGCT ACCAATACAA ATGATGAGGA TAAAACAAAT GACTGAAAAT CGTTATGAAC 540

TAAATAAAAA CTTGGCACAG ATGCTCAAGG GTGGTGTTAT TATGGATGTG CAGAATCCTG 600

AACAGGCTCG TATCGCAGAA GCTGCTGGTG CGGCAGCTGT GATGGCCTTG GAACGAATTC 660

CGGCTGATAT TCGTGCAGCT GGAGGAGTTT CCCGCATGAG CGACCCAAAG ATGATTAAGG 720

AAATCCAAGA AGCGGTTAGT ATTCCAGTAA TGGCTAAGGT CAGAATCGGG CATTTTGTTG 780

AAGCTCAGAT TTTAGAGGCT ATTGAAATTG ATTATATCGA CGAGAGTGAA GTTCTATCTC 840

CAGCTGATGA CCGTTTCCAT GTGGACAAGA AAGAATTCCA AGTTCCTTTT GTCTGTGGTG 900

CTAAGGATTT GGGTGAAGCC TTGCGTCGTA TCGCTGAAGG TGCTTCCATG ATTCGTACCA 960

AAGGAGAACC AGGGACAGGG GATATCGTCC AAGCTGTTCG TCATATGCGT ATGATGAATC 1020

AGGAAATTCG CCGCATTCAA AACTTACGTG AGGACGAGCT TTATGTTGCT GCCAAGGATT 1080

TGCAAGTCCC TGTAGAATTG GTCCAATATG TTCATGAACA TGGAAAATTG CCAGTTGTAA 1140

ATTTCGCTGC TGGAGGTGTT GCAACGCCAG CAGATGCTGC GTTAATGATG CAATTAGGGG 1200

CAGAGGGGGT CTTTGTCGGT TCAGGTATTT TCAAGTCAGG AGATCCTGTT AAACGAGCGA 1260

GTGCCATTGT TAAGGCTGTG ACTAACTTCC GTAATCCTCA AATCCTAGCT CAAATCTCTG 1320

AAGATTTAGG AGAAGCCATG GTTGGTATTA ATGAAAATGA AATCCAAATT CTCATGGCTG 1380

AACGAGGAAA ATAGATGAAA ATCGGAATAT TGGCCTTGCA AGGGGCCTTT GCAGAACATG 1440

CAAAAGTGCT AGATCAATTA GGTGTCGAGA GTGTAGAACT CAGAAATCTA GATGATTTTC 1500

AGCAAGATCA GAGTGACTTG TCGGGTTTGA TTTTGCCTGG TGGTGAGTCT ACAACCATGG 1560

GCAAGCTCTT ACGTGACCAG AACATGCTAC TTCCCATCCG AGAAGCCATT CTATCTGGCT 1620

TACCAGTGTT TGGGACCTGT GCGGGCTTAA TTTTGCTGGC TAAGGAAATC ACTTCTCAGA 1680

AAGAGAGTCA TCTAGGAACT ATGGATATGG TGGTCGAGCG TAATGCTTAT GGGCGCCAAT 1740

TAGGAAGTTT CTACACGGAA GCAGAATGTA AGGGAGTTGG CAAGATTCCA ATGACCTTTA 1800

TCCGTGGTCC GATTATCAGT AGTGTTGGTG AGGGTGTAGA AATTTTAGCA ACAGTGAACA 1860

ATCAAATTGT TGCAGCCCAA GAAAAAAATA TGTTGGTAAG TTCTTTTCAT CCAGAATTGA 1920

CTGATGATGT GCGCTTGCAC CAGTACTTTA TCAATATGTG TAAAGAAAAA AGTTGAGATT 1980

GAATTTCTCA ACTTTTTTAC ATGTAATAAA CAATAGCGAT GTATTGAAGT GCGGACGCAG 2040

CTAGGATAAA GAGATGCCAA ATCATGTGGA AATAAGGTTT TTTCTTGGCA TAAAATCCAG 2100

CTCCAACTGT ATAACAGAGT CCGCCAGTTA CCATGAGACT CCAGAAAACG GGTGTCGTTT 2160

GACTGATAAT GGCAGGAATG ATAGCCAGAA CCAACCAGCC CATAATCAGG TAAAGAGCAA 2220

GGCTAAATTT CTCATTGACC TTTTTAGCAA AGATTTTATA GAGAATACCA AAGATGGTCG 2280

TTCCCCATTG GATGACAATA ATCAGATAGC CAAACCAGTT ATTCATCAAG GTCAAGACAA 2340

CGGGCGTGTA TGAGCCGGCA ATGGCAACGT AAATCATAGA ATGGTCAATG ATTCGCAAAA 2400

CATATTTGTG GGTCGAACCA TAGGCCATAG AGTGATAAAT GGTGGATGAT AGGAACATGA 2460

GAAAGAGACT GATGACGAAA ATGGAAACGC CGATAGAGGA TAAAAATCCG TGTGCTTCAT 2520

AACTATAGAT GGATGAAATA GGCAGCAAGA TAAGCATGAT GACTGCACCC ACAGCATGGG 2580

TCACGCTATT AGCAATCTCC TCTCCAAAAC TGAGTTGTTT GCTGAGTTTA AGACTAGTGT 2640

TCATTGGATT ACCTCCTCTT GAGTATGATC GATTAAGTCT AGAGTTTGAT GATAGAGTTT 2700

AACGGTTTGG CAGCTGGTTT GGATAATAGG GTTAGCTGGG TCAATTCCTT GGTTCATGTA 2760

GTCCACAAAA GCATCGTAGA GTTGGTCTGA ACTTGCTTGA GTTTGTAGAG TATTAAGTGT 2820

CTGGGCTATT TCTTGAATAG AAAATACAGA CTTGAGGGTT GTGATAGCAA TCAAACGGGC 2880

AATCTGTTGG CGTTGGTATT TTTTTTTGTC AGGCTTTGTC AGGTAACCAT TTTTCACATA 2940

ATTGTTGACC ATAGATGCTG TTAGGCCCTT GTCTTTATTA GGAGAGATAG GGGCGCAGAC 3000

CTGATTGACA 3010

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15213 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

CATAAATCGG TGCAAATAAC TTAATAGTGA AGTAGCCATT TCTTTCGTAT TTACCTGAGG 60

CATATTCCCT AGACGAAAGA ATATTATTAT CAATCAAATC ATTGAATGAA CGTAGTCTTT 120

CAACTTCTTC TACTGTTAGA TTTCTGACAA CATTTGTTGC ATAGACCTTA TTTCCATCAG 180

GATCAGGATG GTACTCATTT GTAACTTTTC TAAGAAGTTG TTGTTTTTGA TTCGTATCCA 240

ATTTAAGAAT TGAATTTCCT TCGAGATATT CCAACATATA AACAACGTCA AACATGTTGT 300

GGACATATTG CTTCAAATCA TCTGCATTAT TAAATCTTGT AGTTGGATCA AGTACTTGTA 360

ATCGTCGACT TTCTGTACTA TCAGATTTTG AATGTTTCAA GATGGAGTTG ATGGTAATGG 420

TCGCATCATC TGGATGGTCT GGTGCTTGTA ATAATCCTTT AGCAAAGAAC TCTGGTCCCA 480

AGCCACTTCT TCGACCATAT CCTCCAAGAT AAATGTCCTG ATCTGAGTCA TGTGTCATCT 540

CATGCGTATA AGTAATAGCT CCATCCTTAT CCAACATTCG ATAACCCATA TAATAAACTG 600

CATCACCTGT AGCATAAGCA CCGTGTTGAT TATGCCCAAC TTTATTTCCA ACAGGTCCAA 660

AGAAATGTTG CATTGCAGGA TTTGGATTAT CAAAATCTGC CACTTCTGTA GCTTTCCCTA 720

CGGTATTATC ATCGCCAAAT TTATAAGCAT CGTAAAGCAA AATATTTCTA TAAAGTTTTT 780

CACGTGCATT GTCGTCTAAA ATACGATACC AATAATCGTA GTGATCTCGC TGACGTTTGG 840

CTGTTTCACG CGCATTTTCT TCAACAAAAT CATTGAGAGC CTTGCCCGCT TTATGGTCAC 900

TACTGCGGTA GCGATCATAA GCTCCAAATC CTAGACTAGA CATGGTCGAG ATGACAAATA 960

CGGATCTCTC TGGCAAGGTC AGGAGAGGCA AGACCATATT GCGGTATTTC CATGTGGCAC 1020

TCGTGATACG ATCATAAACA CCGATAGAAT ACTTGGTGCC AGCTAACCCT TGCTTCGTTT 1080

TCACCTCTTC GATAGTGGAT TTTTCTTCGA CAATGTAAGC CTTAGTCTCT GATTTAAACC 1140

AGTCATTATT GCTTGTATTT GGTAAAAAGA CTTTTCGGTA ATGTTCCAGC GTGCTAAACA 1200

AATCTGTCGT TCCATGTTGA CTGGCAAGAC TGATACCATA AGTATCGACA TTATTCTTAG 1260

CTAGAAGATT GTTAAAGCCA GATTTACCCA ACTCAATCAG AGTATCTAAT GGTGAAGCAT 1320

TCCCCTTACC AAAGAAGTCC AAATGGTACA GAACTAGGTC TTTGACATTC ACCTGACCAT 1380

AGCTAAAGTT ATACCACCGT TCCAGATAGG TCAAGCCAAG TAGCAAGGCT TCCTTGTTGC 1440

GTTTGATTTT ATCTACAAGA TAACCTTCAG TGACGGGGTT AGCACTAGCC AGTCCAGCAT 1500

CCGCTGACAA GAGTTTTTTC AAACTGTCTT CCAGTTGTTG TTTTGTTTTG GCGAACTGGT 1560

CTTCTAGATA GAGCTCAGTT TGCTTGACGT TTGGAGAAAT ACCCAGCGTC TTTCTGATGG 1620

CTTCTGAATG ATAGTCAACC TTTTGTAAGT CAGGTAAGAC TTGCTTGATG ATAGAGGTTT 1680

GGTCATACAG GAATTGGTTT GGCGTATAGA GAAGTCCAGT ATTGCCCAGA CTATATTCTG 1740

CTAATTTGGC GAAATCATTC TGGTATTTGA GATCCAGCTT CTCAGATAAA TCATCCTTGT 1800

AGTGAAGCAA GAGTTTGTTT GCAGTCTGTT TGTTAGAAAC AATGTCTGTG ATGACTTGGT 1860

TGTCCTTCAT CATGACTGCT GACAAGAGTT CTTTTTGATA TAAAAGACTG TTCTCATTGA 1920

CCAGGTTTCC GTATTTGACG ATGGTTGCCT TGTTGTAGAA AGGTAGCAAT TTTTCAATGT 1980

TTTTATAAGT CAAGTTGCGC TTAGCTTGAT AATAGGCCAC CTTAGAAAAA TCACTGTCTT 2040

TTTTGCCACT TGTTGAAAGT GGCTCCACTG TTGGTAAAAT GAGAGGATTG ATTTCTGCTT 2100

TTTTGCTTGC AATTTGAGAA GCATCTAGCA TTGTTCCTCT TTCTTCAAAG GATTCCTTGC 2160

TGACGACCTC ATCCTTGACC AAGGTGACAT TGTAGACTCT GTTGGCCTTG CTGCTGAATG 2220

TGTCCTTTAC CTTCATTTCG TTATAGTGGT AACCAGTGAT GGCATTTCCG TTGGTTACAT 2280

TAACATCGCT GAGAACATTG GTCAAACTTC CAGCATGCCT AACATCACCA GAAGTTCGAT 2340

CCCACAAATT GCCTGCCACT CCAGCGACTC TACCAAAGTG CTTGACATTG TTGATATCAC 2400

CTTCAGCATA GCTATCTTGG ATCTGTGCAT CTCGGTCTAC TAGGCCTGCA AGTCCACCCA 2460

CAGTCTGATC TGAAGTATTT GTGTTAGATG AAATGGCTAC TGTCGCTTTT GACTTAGTAA 2520

GTAAAGCCTT GTCACCTGTC AAATGACCGA CCATACCACC GATATTGTAG GCAGCAGTCG 2580

TTTCATAAGT GTTGATAATT CTTCCCTTGA AACTGCTCTC TGTGATGCTT GATTGCTCAG 2640

CCTTAGCCAG CAAACCACCG ATACCACGTT CACCAGCCAG AACACCATCG ACGTGAACTT 2700

GCTTAATTTT TGTGTTATTC TGAGCTTCAT TTGCCAGTGA ACCGATATCA TCTTTCCCTG 2760

AAATAGCAAC ATTTTTTAGA CTCAGTTTTT CTACTGTAGC ACCACTCAAG TTTTCAAACA 2820

GAGGTTTTTT CAAATTATAG ATAGCATAAT TCTTGCCATC TTTTTCACCG ATTAAACGAC 2880

CAGTAAAGGT GTCCTTGATA TAGGATCTTT CATCAGGACC AAGCTCCACT TCGTTAGCAT 2940

TCAGGCTGGC CGCTAAATGA TAGGTTCCAG AGGGATTTTG GTTTATAGCT TTGACCAGAT 3000

TACTAAAGGA AGTAAAGTTT GTTGTTTCTT CTGTTCCCTT CTTAGCTAGA TAGAAGGTAA 3060

AATTATCTTT ATATCTGCTT TCTATCTCCT GCTGAAGCTT CTCTACTTTT GCTGTGATTT 3120

TATAAAGGAT TTTATCATTT TTTCTTTCCT CTGATATTGA TGCTACTGGT AGGTATACAT 3180

CTTTGAATGA AGAAGATTTC ACTTTAACAA AGTAGCTATT TGGATTGCTT GGAACTTGCT 3240

CTAACGAAAT GTGTTGTTTA TAAGTACCAT TTGACAAACT GTATAACTCT AGGTCGGAAA 3300

CATTTCTTAA TTCAAGTGTT TTCTCTGGTT CTTCTACCTT TTTATCAGGG TCTAGTTCAT 3360

TTTCTTGTTT AATTTCTTCG TTTCCATTTG AATTGGATGT GTTTGATTCG GTTGAAACAT 3420

CCTCAGTTGA ATTTCCGTTT GATGGTTCTG GTTCTGTTTG TCCATTCTCT GATGTTGTAT 3480

TACCTGAATT TTCTGGTTTT GTTGCAGTTC CGTTTTTTTC TGGTTGATTT GATTCTTCAA 3540

CTGGTGGTTT TGAATCACTA GGTTTATTGG ATACTTCTCC AGTATTTTCG TTAGCTATTT 3600

TCCCAGAGTT TGTTTGTGTT TCTTCTGCAG GTTGAACTGG TTTTTCTGTT TCTTGATTTG 3660

AGGTACCTTC TACTGTGCCT TCATTTGGAT TTACTGGAAC TTCTTCTACA GTTTTTTCTG 3720

AATTTTCATT TTTAGAGTCA TTATGTTCTG GTTTATTTGA TTCTCCAACT GAGGTTGTCG 3780

AATCACTAGG ATTACTGGAC ACTTCCCCAG TATTTTTGCT AGATGTATCT GGTGATACTT 3840

TCTCTGAATT CGTTGTTGAT TCTTCTGCAG GTTGAACTGG ATTTTCTGCT TCTTGAATTG 3900

AGGTTCCTTC TGTAGTACCT TCATTTGGAT TTACTGGTGT TTCTTCTGTT GGTTTTACTG 3960

GAACTTCTTC AGTTTTTTCT GGACCTTGTT CTTTGGTCTT CTCAACCGGA GTTTCAGGTT 4020

TTACTTGCTC AATATTACCC TTATATTCTG GAAGCGGTGC TACCTGCTCT GGTTCACCTT 4080

TATCACTTAC CACAGTATCT GGCGACTCTG GTTGAACCTC AGTCTCACCT TTGTCGGTCA 4140

CAACTGCTTC GGGTAATGTA GGTTGAACTT CTGGTTCGCC TTTGTCACTT ACTACAGCTT 4200

CGGGCAACTC AGGCTGAATT GCGGGTTCAA CAATAGCTCC AGACTGTACG TCCTTATGTT 4260

CTACACCAGT CTCAGGTTGT TCCTTTATAA CTTGAGTTTT TTTAGTACCT TTTTCGACTA 4320

TTCTTGGACT AGGCGCAGTC GTTGAAGTTG AAACAATTTC TCGCGAAACT TCTTCCTTGT 4380

TTACAGAGAA TATTCTGACG ATTTCAACTT TCTTACCTAA TTTACCTTCT TGTTTTACTC 4440

TTACAGTTCC TTCAGCTAAA TCAGGATTTT CTTGAATTTC TTCTTGAAAA TCTATTTTTG 4500

TCTCCATAGT TTCCTCACGA TATAAGAGTT CAGGTTTGTT CAATTGACCT GATAAAACTT 4560

CATCCTGTGG ATTTAATGTA TTTACCCCAG TCTTTTCTTT TGGAGAAATC TTCTCCTCTT 4620

TCTTCGTTTC TAGATTCTTA TGTTCGGCTA ATTGTTCTTG AGAATCTGAA GATTGTTTCT 4680

CTTCTTTTCT TGGATTGATT AATTCAGTAG AGAAAGGTTT TTCAACTACT TGAACTTCTG 4740

TCGGCTTAGT TGAAGAAACA GGTGTTTGTT CCTGAATAGC TTGTACTGTT GATGGATGGT 4800

CTACAAAATT CGGTGTAACA TTATAATCCA CCTTTTGTTG TTTTGTAGGA GTGGCAACTG 4860

AACTCTTTTG ATTACTTACT TCAGACTCAG AAGTCGTTTT TCCCTCTTTG ATATATCCAA 4920

TATAAGTGTA ACCTGAAATC TCTTTAGGAA GAGGTAATTT TTCTCCAGAG GTCAATTCAT 4980

AGTCCGTATT GTAATTTAGC AAAAGATGAT TTTCTAAAGC ATGGACTGAA ACTAAGACAC 5040

CATTTCCTAT CCCTGCAACC AATACTAAAT GTAATACCGT TTTATTCTTA ACCTTTTTCT 5100

TGGAAACAGC AAAAATTAAA ATTCCCATAG CAGCTAAGCT AGCACCAGCA ACTAGGGCTT 5160

GCCTCTCATT CTTGCTTCCA GTATTTGGCA ATTCCGCCAG TTGATTTTGA GAATTTAACT 5220

TATAAACAAG ATAATAAGTT TCATCATCAT TCTCCACGTA TGTCGGAATA TCATAGACAA 5280

GCTGCTTCTT TTCTTCTGAT GATAGCTCTG AATCTGCCAC ATATTTATAG TGAACTCCCG 5340

CAGTTTCTTG AGCATCCACA GATGAACTAG CTAATACAGA CATAAAAAAT AAACTTGAAA 5400

TCGTTGCAGA TACAAGTCCT ACTGATAATT TTCTAAATGA AAAACGCTCT TGTTTTTCAC 5460

CAAAATACTT TTCCATTATT CCTCCTTGAA ATAAAATTTA TATATGTTAC AAAGACCTTT 5520

ATTATATTAG TGTATTATCT ATTATCTATA GAAAAGGCAG TATACCTTAA TTATACTCTT 5580

AATTTACAAA AAAGTCTTAA AATTGAGATG CGCTTTCATA CTTTGTTTTA TATTATTTGG 5640

AGGTACAATA ACACCTACCA TGAAATTTAC ACGGTAGGTG TTACTCATAT CACTAATCGT 5700

300

TCTAAAAATG GTTTGAGGCA GTTGAGGAGA ATTCCTTCTA TCCAGCTTCC TTGTGCTGAT 5760

GAGCGATGGT CTTCCTGCAG GCTTTTTTTT AGAAAATCTC GGACTTGTTC TGGTGCGATT 5820

TCAAATTCAA AGGCTTTCAT TTTATAGAAA AAGTCGATGA GATGATCTGA CAGGTATTCA 5880

GTTGAAAAGG GTACTTCACC ACTTTTTCTA TATTCTAATA AGAGTCTAGA AAATCGAGCT 5940

TTTTCTTCAG GAAGCTCACG AAAATAGGAA TTGAGGATCC AAGTCTGCTT CTGTTTTCTT 6000

TCAATTGGAT CCTGACTGGC AATTCGTTGG TCTTTTTCCA GCTCTTTTTG GTATTGTTTG 6060

GCCTTGATAG CTCGTTCTGC TCTATTTTTA CCAAAAAGAA TTTTTTCCCA CTTGCGTTCT 6120

TCTTGAGTCA GGGTCTCTGT AAAGCCAAAG TAATCTTGAT AAGCACGCTC TGCGGGTCCC 6180

ATGGCTAGAA CCAGATTGTC TGCATATTGC TTGGCGATTT TATCCCTCTT CTTGCGTTCT 6240

TTCTCTGCCT GGATACGGAG TTCTTGTTCG TAGTCAATTT TCTCCTTGCC TAGCTTGACA 6300

AGGTAGAGTT GGTCATCCGA TTTCCCAAGT AAAAAGGGTT TGATACACTT TTCAAGGACT 6360

TCTTCCATCC GAGCCTTTTT CTTTGGTTCC GCCTTGGTCC AACTTCCTCC CTGAAAGACT 6420

TCTAGGAAAA GCTGGTAGTC TCTCTCAGGC GCAAATTGAT TGCCACGATT GGGTTTGAAA 6480

ACACCTTTTT CCCAGAGCCA TTTTAGAAGT CGCTCGTCAA AGTTACTTTT ATTGACCTTG 6540

ATTTTTTCCT TTTTCTGAGC TTTTCTGGTT AGATTTTCAA CCTTTCTGAG CAGTTTTTCT 6600

TCCTCTTCCA ATTGCTGGTC AAGGGACAAT CGATGAAAAT GACGAACACA GTCGCTACCA 6660

ATTGGAAAGA GGCGTTGGCC TGTGACACCG TTAAAGAGTT CATAAGCGTA TTTGATGGCA 6720

TTTCCACAGA CACAATTGCT ACGGCCGATA CCGTTAAAAA TAAAGGAAAC TTCATTCCAT 6780

TCCTTGGTAG CTTGTTCCCA AGTATCCGCT TTCGAAGCCT GTAAAACTGC ATCGTGCAGG 6840

GATTTTCTAA CTGGAAGTGT CATGAGGTCT CCTTTCTAAT ACTCAATAAA AATCAAAGAG 6900

CAAACTAGAA AGCTAGCCGC AATCAGCTCA AAACACTGTT TTGAGGTTGT AGATAGAACT 6960

GACGAAGTCA GCtCAAAACA CTGTTTTGAG GTTGTGGATA GAACTGACGA AGTCAgTAAC 7020

CATATATACA GCAAGGCGAA GCTGACGTGG TTTGAAGAGA TTTTCAAAGA GTATAAGTTA 7080

TACTTTTACA ACTTGAACCT CGTCTTTACC GAGTAAAATC AAGTATTTTT CAATATTTTC 7140

AATCGAATAG GCTCGTGATA AAGCCTCTTC GTATAGAGCT AACTGACCAC GATAGCGGTC 7200

TACGAGTTGA CTTGGTTCAT CATAGCGGTC TGTCTTGTAG TCGAACAGAA CAATTTTGTT 7260

TTCGTAAAGC AGATAGCCAT CAAGGATACC ACGGACAACA AAGTCTTCCT GACTCTTTTG 7320

GTCTCGTTTG AGCATGGAGA AAGGTTGCTC GCGATAAAGA TGGTCGGTAT TAGCAAGAAT 7380

TTCCTGACCG AGTACTGTGT CAAAGAAAGC AAGAATTTTA TCAAGATTGA TCTTGTCTCT 7440

GACAGCTTGG CTAGTTTGAA CTTGTTTGAG TGTTTCTGTT AGGCTAGCAA GGGTTAGTTG 7500

CTGGCTGAGG TCAATTCTCT GCATGAGTTC GTGAGTAGCA CTACCAATCT CAGCTCCAGT 7560

TACCTTTTCT TTGGTTGAAA AATCTGGCAA ATCGAAGCTG ATTTTCTTGC CTACTGACTG 7620

ACCTTGACCA GCAATCTCGA CACCTTCCAT ATCCATAACT GGTTCGTAGA ATTTCTTGAT 7680

TTGACTTGGG GTTTGAACAC TAGGAAGTTC AATAGCTGCG CGGTGAAGAG TATTATAAAC 7740

TTCCACCTCC TTCAGCATTT CCAGAGCTTC TTTGATGGTA TCTGACTGAC GATTGTCTGC 7800

TTGGGAGCTA TCTTGGAGAG GACTCTTGGT TTCCAACTCT CCGATAGCTT CTCTGGTCAA 7860

CTGATCTTCG CCAATAAAAC GATAACTAAA GTTGAGCTTG TCCTTAGTAA ACACTTTACT 7920

GATAGCCCAA AGCCAATCTT GGAAATTCCG TGCTTGCAGT CTAGTATTGC TATTTAGTTT 7980

CCCATTTTTG GCTGCTGGGT ATTCCTTGGA TTCCAGCTTT TCACGAGAAC CCTTGCCGAC 8040

AAGATAGAGC TTTTTCTCAG CCCGCGTCAT AGCAACATAC AGCAAACGCA TCTGCTCAGA 8100

ATAGCTTGCT AGCTGTAATT CCTCTTCGTT CTGCCTATAG GTCAGACTAG GAATGGAGAG 8160

TTTGATGGTT TTAGGATAGT GGTCTTCTAC TGCCCCTGTC TCCATCTTGG CAATATATTT 8220

GACACCAAGA CCATTCTGAC GACTGAGAAT GACTTCTGAC ATAGAGTCTT GCTTGTTGAA 8280

ATCTTGATCC ATATTGAGGA TAAAGACGTA AGGAAACTCC AGCCCTTTAC TCTTGTGGAT 8340

GGTCATGAGC TCTACTGCAT CTTTTGGCGG TGCGACGGCC ACGCTTGCCA AATCGTGCTG 8400

GGCTTCTAAG ACTTGGTCAA TCATACGAAT AAAACGCGAC AAACCTTTGA AATTGCTCTT 8460

TTCAAATTGA TCAGCACGCA GTGCTAGGGC ATAGAGATTG GCCTGCCTAG CAGGACCATT 8520

CGGCAAAGCC CCAACATAGT CATAATAAAA ACGGTCGTTG TAAATCTTCC AAATCAAGTC 8580

ATAGAGAGAG TGGGTTTTGG CATACAAGCG CCAAGAAGCT AGGATATCCA TGAATTGCTT 8640

TAGTTTTTCA GCTAGAGCTG TGTGAATCAA GCCTTTTTGA CTACTTGCCA TTTTTTGTGC 8700

ATTGACCAGT TTCTCATAGA GATTTTCGTG [illegible] 8760

ACGTGCTAGC TCATCCTCAT CAAAACCAAA CATTGGAGAC TTCATAAGGG CAACCAAGGC 8820

GTAGTCTTGC AGGGGATTGT GAATGACACG AAGAGTGTCT AGCATGACTT GCACTTCTAG 8880

GGATTGGAGA TAATTGTTTT GCTCTCCGTC AGTTTTGACA GGAATTCCGT ACTCAGACAG 8940

GGCGAGGAGA ATCTGGTCAT TACGACTGCG GCTGGAGGTC AGAAGGGCAA TTTCCTTAAA 9000

GGCAACACCT TTTTCTTGAT GAAGTTTCAG AATCTCCTTG ATAACTAAGC GCATTTCGCC 9060

TGTTAGTTTC GTTTCTGTTT GACTCTCTTC TTCCTCACCT GTATCGTCCT TGTCGTAGAG 9120

GAGAAATGCT GCCTTGTTGT CTGGATTGGG AGTCAGTTTG GTATTGGCAA AAACAAGCTG 9180

GTGCTTGTTA TCATAGTTGA TTTCGCCGAC CTCTTGGTCC ATGAGACGTT CAAAGACATC 9240

ATTGGTTGCT GACAGCACTT CTGAACTACT ACGGAAATTT TCCTTGAGGA TAATGAGCCT 9300

GCCTTCTTGG GGATTTTGCG CATAGCGTTG GAATTTCTCA TTGAAAATCT GCGGGTCTGC 9360

CTGACGGAAA CGATAGATGG ATTGCTTGAT ATCTCCCACC ATAAAGCGAT TGTGGCCATT 9420

AGACAACAAT TCCAGCATCC GTTCTTGAAT ATGGTTGGTA TCCTGATACT CATCGACCAT 9480

GACTTCATGG AAGCGCTCCT GATAAGACTC ACGAACTTGT GGGAAATTCT CTAAAATCTC 9540

AATGGTGTAA TGGCTGATAT CAGCGAATTC GAAGGCATTT TCCTGTCGTT TTCTCTGACG 9600

ATAAGCCTCT ACAAAATCGC TCATGAAAGA TTGGAAGGTT TTAGCTAGTT TCCAAGTGTC 9660

TCCATGATAA CGTTCTTGAT AGTCGAGAAT CGCTATCTGG TCTGATAATT GTCCTAGTTT 9720

AGCAAACTGG GTCTTTCTCT CTTCGTTGTA GGCATCAGCC AGGGGCTTCA AATCAGCCTA 9780

CGGCTGGCAT TAGTCAGAGC TCGACCGTTT TTCTCCTTAG AGATGGCGAC AACACGCGCA 9840

AGCACTGCCT GATAAGCCTG ACTATCGGAC TCCTGATTTA GGGAGCCAAT TTCATCCAGA 9900

ATTAACTGAA CATTTTCTAA ATAGGCAGCC TTTGCAAACT CCTTGGCATC GTTATCCAGA 9960

TGGTAACGGA AAAAGCTTTC CAAATCCCAA AGGGCTTGTT TGATTTGCTC GGTCAGTTTT 10020

TCTTTTTCAC TGGTAAAATC AGCTTTCTCA AATCCTTTGA GGAAAGATTC ACTCAGCCAC 10080

TTTTGAGGAT TACTGGTGGA TTGGAGGAAG TCATAGATTT TATAGACCTG CTGGCGCAGA 10140

CCCCGTTCGT CCTTGCCACG CCCAGCAAAG TTTTTCAGCA AATGACTAAA GGTCTCTTTC 10200

TGTTTACCTT GGTAATGCGC TTCAAAGACC TCATGAAAGA CTTCGTTTTC GAGAATAAGT 10260

TGCTCGCTTT GGTTTTGTAA AATACGGAAA TTAGGTGCAA TATCAAGCAG ATAACCATGT 10320

TTGCCAAGGA ATTTTTGTGT GAAAGAATCC ATGGTTCCAA TGGCAGCGTT GGGTAGGTCT 10380

GCCAACTGGC GACCCAAGTG TTGTTTGAGG TCGACATCAT CTGTTTCTTG GATTTTCTTG 10440

CTGATTTTTT TCTCTAAACG TTCTTTAAGT TCAGTTGCAG CCTTGACGGT AAAGGTTGAG 10500

ATAAAGAGTT GAGAAATTTC GACACCACGC GCCAATTGGT CCAGAATGCG CTCTGCCATG 10560

ACAAAGGTCT TTCCAGAACC AGCCGATGCT GAGACCAGGA TATTCTGGGC AGAAGTGTAG 10620

ATAGCTTCGA TTTGCTCGGC AGTTTTCTTC TGTTCCTTGC TCGAATTTGC TTCTGCTTCT 10680

TGCAGTTTTT GAATCTCCTC CTCACTTAAA AAGGGAATAA GCTTCATCGA TTCAACTCCT 10740

CTCTTATTTT TTCAAGCCAA GCTTGCTTGA GTTTTTCTCC GACCAGACGC TTGCCATCAG 10800

CTAGGTCCAA CTTTTCTAGG AAACGGGCTT GGCCCAGATG GTAATTGGCT TCAAAGCCTG 10860

TAATAGCCTG ATGTTGCTGG ACGTATGGGG CAATGCTTCT GCCATTTTCA GTATAAGGAT 10920

TGATGGCGAA CCGGCCTGCT AAAATCTTCT CAGCAGCTTT CTTGTAAAGA TAGGCATTGT 10980

AGTCCAGTAG GAGCTGAAAT TCCTCATCTG TCAGTTGATT AGCCTTGTTT TTGTTATAAA 11040

ATTCGCCTAA ATAACTGCTT TCTTTTTCCA AGAAGAGCCC TTGGTATTTC ATAGATTTGC 11100

TGGCTTCTAC CACTGCTCCT GCCAGACTTT TTACCGCCAT CAGAGATTGG ACAGGTTCAG 11160

CCATTTCCAA GTACATGGCG CCGAAAAAGT TCTGCTCCCC TTCTCTTTTT AGGGCAGCAA 11220

GATAGGTTGG TAACTGAGAA TTGAGCCCAT TAAAGAAATG AGGAAACTGG AACTGAGTCA 11280

GACTGGATTT GTAGTCTACT ACTCCTATCG CTCCATTAGC TTTCAAACGG TCAATCCGGT 11340

CCACCTTGCC TCGTACAAAG ACACTGCGTC CATTGTCTAA TTGAATAAAG GCTTGGTCTT 11400

TTCCACCAAA ATTTGCTTCT TCTTTGATGG TTTCGATGGC TGGATTGTGT CGGAGAATAT 11460

GTCCAGTTGT CCGTGCAACA TCAAGCAAAA CTTCCTTGGT AAACTGGGCT TCCAAACTTT 11520

CTTGATAAAT AGCTTCAAAT TCGCGTTCTT GACTGGTTTC TTGAATAGCT TGTTCTAGAC 11580

GTTGGTCAAA GGAATCTTCA TTAGGCAACT GTAAGGCGCG TTCAAAGATA CGATGCAAGA 11640

AATTCCCGTG ACTACGGGCA TCAGGATGCA AACGTAATTC CTCCTGCAAG CCTAAAACGT 11700

AGCGTAGGAA ATAACTGTAT TCATTGCGAT AAAACTCTGT CAAACCCGAC GTAGACAGGT 11760

AAAACTCCTG TTTGGCAGGA TAGAGAGCTT GCAAGGTGTC CTTGGCTAAG GTCTTGCTGC 11820

TTGGACTGGT TGGGATAGCT GGATTTTCCA GACCTTGCTG ATCTAGTTTT TTACCTATGA 11880

CACGCGACAG AACCTTGACA AAAGTCAAAT CTTGCTCAGT ATCGCTCATC TCACCCTGCT 11940

GGTGATAGGC AACCAGACTA GACAAAAGAC TGTGATAGGA CCCCATATCC TCCTTAGACA 12000

GTCCTTTGTG ATTCATCCTC TTCTCTCTCC GCCTAAATCC AAAATGGATC AACTCTTGAA 12060

GATAGGCAGA TTCCTTACTT TCACTTTCGT TAAAAAGGCT TGGAGCCGAC AAGAACAACT 12120

GCTTACGAGC AGAATTGACC AAGGAAAGCA TAGTGTAGCG ATTTTTTCTTG AGATTTTCAC 12180

TGCTGGCAAT CAGTAATTGA ACGCGTTTCTT CGGTCGCTTG GTTTATGTTT TGCGTTTCTT 12240

CATCTGTCAG AAGACTGGTG TTTTGAGAAA TTTTTGGTAA ATTGTCCTGA GTTAGTCCAA 12300

TAGCATAGAC AAAGTCAGCA GTCAATGGTG CAATCAAATC GTAACTCTGC ACCAGAACAG 12360

TGTCCACTGT TGCTGGAATG GTACGGTATT GGGACAAACT CATTCCAGAA TGGAGCAAGG 12420

CTAGGAAGTC TTCCAGACTA ACCTGTGAAC CAGCAAAAAC AGTCGCAAAT TGTTCTAAAA 12480

CATGGCAGAA AGCCTTCCAA ACTTCGGCTT GTCTTTCCTG TTCTACAGCT TCCAAAGTGG 12540

TTGTCAAATC TTGTAACTGC TTGGTCACAG CTCCTTCTTT TAGAAAGACA CTCCATTTTT 12600

GTAGGAGTTT TTCAGCCTTT TGTTTTCGGC TGGCAAAGAG GGTTTCAAGA GGTGCTAAAA 12660

TTCTCAGGCG GAGGACATTC AAACGCTCAA GATTAAATTT TCCATGGTGG GATTTGGTGA 12720

AGGTTTGCTG AAAGGCTGGC AAGCCATTGA TACCAAGATA GCGGATATAT TGCTCAAAAG 12780

CATCAATATC AGACTGACTG AGGTCAGTAT ACAAATCAGT TGTAAGAAGA TTAATCAAAT 12840

CCTCCTGACG AAAACGGTAA CGTTTTAAAG CTAAAATAGA CTCGACAAAC TGAGTCAAGG 12900

GATGATGAGC CATGGCTTCG CTTCTACCAA GATAAAAAGG AATCTGATAC TGGTCAAAAA 12960

TGGTTTTGAG AGATAACTGG TAAGAAGCTA CATCCCCCAA GAGAATACGA AAATGCTTGT 13020

AGCTCAGGTC TGAGTTCTCA TGTAATTTCT GACGAATACT ACGGGCTACT AGCTCCAACT 13080

CCTCCTTTTG CGTCAAACAA GACCAGATTT GTAAATTTTC ACGGTCTTTC TCATCGACAT 13140

CCAAAGCGAG TTCTGAAAAG TCATAAGAAG ACTCCAACAA ACGAGAGGCC TTGTCAAAAC 13200

TATCCATCTT CTCATGAGTT TGAGAACAGT CCTGAGCAGG CGTTTGGTAT TTAGAAGCCA 13260

GATGATGGAG AAATTTTACG CTGGCTTGGT AGAGATTGCC CTCGCTAAAA GGACTGGTAT 13320

AGGCTTTCTT ACTAGCATAA GCCCCGATAA CAATCTCAAC ACCTTTGCCG TGAAGTAAGT 13380

CCACAACCCG CTCTTCCTCA GCAGAAAAAC GAGTAAAGCC GTCAATGACC AAGGCGATTT 13440

GATTAAAATC ACTACTTACC TTGTCATTCT CAATAGCCTC AATCAAATGG GACAACTGAC 13500

TTTCCTGGGC TAACTGACCT TGATTAAGAT AGGCTGTTAC TTTCTCAAAA ATCAAGAGTA 13560

AATCCGCCCT CTTATCCTCA TCTGTTAAAT TCTCCAAGTC CAAAAAACTC ATCTGAGATT 13620

TGGTCATCTC ATGGTAAAGC TCAATTAACT GCTGGATCAA TTGAGGATCC TGCTTAATAG 13680

CGCCATAAAC ACGCAAGTCC TTGGGATCGA GTTCGGCAAG GCATTTGTAA AAGGCCAACC 13740

CAAGACCGAT ATCATCAAGA GTAGTTTTAG CTGGTAAATC ATTCAAGACC AGATAGCGAG 13800

CCATTTGAGC AAAGCGCGTG ACGGTAATCG AAAAAGAAGC CTGCTGGGAC AAGTATTCCA 13860

GCACGGCGCG TTCCTTTTCA AAAGAAAGAG AGTTGGGGGC AATGTAGAAG ACCCGCTTGC 13920

CAGCTGCAAC TAGCTCTTCT GCCTCTCTTG TTAGAATTTC TGTCAAAGAA GTCCGAATAT 13980

CAGTATAAAG TAATTTCATC TCAGCCTCGT TGGAATTTTT CATCACCCTA TATTATACCA 14040

TGATTAGCCT CGTAAATCTG TTAAAATATT TAGGCCATCC TTTCTTTTCT TCATCATCTG 14100

CTAAATCTTA AATACTTAGC TTTACTTGTA TTAGATAGAA TAAGTCTGGC TACTGAAAAT 14160

CACATAATAA AAAAGCCTCG GTAACAAGGC TTTGAGTTTT ATGATTGTTT CTTAGGTACG 14220

GAATACACTT CAATGTGTTG TCCCAGTATC TTAATGTCGA CTGGTAGATT GTCTGATTTA 14280

TCGCCATCAA CATCGGACTC TAATTCGATA TCAGAAGAAG TTTTAATATT ACGTGCCTTT 14340

ATATATTCAA TATTCTTGAT AGAATGATTG AACTATAGTA AATTGAAACT ATAATAGTAC 14400

ACCGTGGATG CTAAAATATT TCTAGAAATT AATTTGATTT CCCTAATCAA GCTATTCGTA 14460

TCTTATTTCA ATCTACTATA ATAAAATGAA CCAAAAATAG TACACAATGT GGTATAATCT 14520

TCTTATGGCA TATTCAATAG ATTTTCGTAA AAAAGTTCTC TCTTATTGTG AGCGAACAGG 14580

TAGTATAACA GAAGCATCAC ACGTTTTCCA AATCTCACGT AATACCATTT ATGGCTGGTT 14640

AAAGCTAAAA GAGAAAACAG GAGAGCTAAA CCACCAAGTA AAAGGAACAA AACCAAGAAA 14700

AGTTGATAGA GATAGACTTA AAAACTATCT TACTGACAAT CCAGATGCTT ATTTGACTGA 14760

AATAGCTTCT GACTTTGGCT GTCATCCAAC TACCATCCAC TATGCGCTCA AAGCTATGGG 14820

CTACACTCGA AAAAAAGAAC CACACCTACT ATGAACAAGA CCCAGAAAAA GTAGCCTTAT 14880

TTCTTAAGAA TTTTAATAGT TTAAAGCACC TAGCACCTGT TTAGATTGAC GAAACAGGAT 14940

TCGATACTTA TTTTTATCGA GAATATGGTC GCTCATTAAA AGGTCAGTTA ATAAGAGGCA 15000

AAGTATCTGG AAGAAGATAT CAGAGGATTT CTTTGGTTGC AGGTCTAACA AATGGTGAAT 15060

TAATCGCTCC AATGACTTAC GAAGAGACGA TGACGAGCGA CTTTTTTGAA GCTTGGTTTC 15120

AGAAGTTTCT CTTACCAACA TTAACCACAC CATCGGTTAT TATAGTAAAA TGAAATAAGA 15180

ATAGGGGGGG GGGGGGAGGG GGGGGGAGGG AGA 15213

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6004 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

TTATTACCTG AAACATTAAA TTTAATTGGA CATCCCGTTA TCAATTTTAT AATATCATCA 60

AGATTTTTAT TATCTGATTC AGGAATTTTA TCTGATATAA CAACACCATT TTCAAGATAG 120

TTCATTAAAT TATTTGATTC ACTAACATTA GTGTTTTGAT CTCCATCAAG CCAAAAATAA 180

TGGTTATCGG AATCTAAATA CGATGAGTTT AAAATATTAT TACAAATTAT TTGATTTGCT 240

CCACCAGGAA TATATCTCAC TACTAAATTC TGTTTAAGAT TCTCACTACC TGAATGAGTG 300

ATAACAAACT CTAGAATATA TTTAGCTAGT CTATCTTCAA CATAAATCAT CTTCCTAGAA 360

TGATACACAT CACCTAATTC AAAAAATGCA TCCTGATAAT CAATATTTTC AATAACATCT 420

ACCTTTTCTC CGTTTTTCAC TAAAAGTTTC ACGGCTTCTC TAGGAAAATC TTTTATAAGT 480

TGTGTAGAAT GTGTAGTGAT AATAATTTGA TGTTTTTTAT TTAAACACTC TTGAAGTAAA 540

AACTCTTTAA ATTTATAGAT TGCACTCGGA TGAAGTGAGA TTTCAGGTTC ATCTATTAAT 600

ATTAATGAAT TTGATTGCGC ATTTACTATA TCATTTACTA ACAAAATAAT TCTAGCCTCA 660

CCTGTTCCTG CAAAAGCCTC GGAATATTCT TTTCCAGATT TTTTCATCCA AATAGTTTTG 720

GAAGCTTTTA TATCATCACC TTTTGAATAC AACTTATGTG TTAAAATTTG AATGTCTGTA 780

TAAGATTCAT CCATTATTTC ACTAATAATT TCACAAACTT TATCATCAAC TTTAACATTA 840

TCTATAACCA TTTCCTTTTT ATAACGCGTA TAGCTACTTG TATTATTCTT TAAAATATCA 900

GCAACTGGCT TAGATCGTAA TCTTATAAAA TCTTGTTTAC TACGTTGAGT AGAAATTTTT 960

TTAAAATTAT AGTGATAGAA AAATAAATCA AAAGCAGAAA CATATTCTTT ACAATCACAA 1020

AAGACAACAT TTTTTTCAAT GCCATCCCAT CTGTCTGTCG AAGAACTTCC AATATATTTA 1080

TTTTTGGGTA ATCTTTCCAT CTCATATTGT TTTTGAGGAG CATATGGTTC CCAATAATCT 1140

AATCCTTTTT TTGTTCCAGA ACGGCCTTTA AGAACTTCTA CATTTCTAGA AGCTTTAATG 1200

TTATAATATG AATAGATTAA ACATTGTTTC CCATCCACTT CATCTATTTG ATCAACATTT 1260

GTACTAAACC AATATTCAGA CACACTTTTA TTGGCTGGAG AACCATATAA AGCTTGTAAA 1320

ATTGAAGTTT TATTTACTCC ATATCTATTA CAGACACCTC AGGATTATTT AACTTATAAG 1380

TTTTAACAGC TACGGAATCA ATTTCAACAG CAACTTGAAC ATCTATGCCT GATTTTTTAA 1440

GGCCACTTGT AGTGCCACCT GCACCGTTAA ATAAATCAAT AGCAACAATT TTCCCCATAG 1500

TATTCTCCTA AAGTTTCTCC TTTTTATTAT AACATTATCA AATGTAAAAC CCAACCCGAT 1560

AGGGTTAGGT TTTTAACATC ATTTCACCAA CTTCTTCATC TCATCAATAC GTGCGACGGT 1620

CGCGTCATAT TTAGCTTGGT AGTCAGCTTG TTTGTCGCAT TCTTTTTGGA CGACTTCTGG 1680

TTTGGCGTTG GCTACGAAGC GTTCGTTAGA GAGTTTCTTA CCAACCATGT CCAGTTCTTT 1740

TTGCCATTTA GCAAGTTCCT TGTCGAGACG GGCCAGTTCT TCTTCAACAT TGAGGAGATC 1800

GGCCAGTGGC AGGTAGATTT CTGCTCCTGT GATGACACTT GACATAGCCA GTTCAGGTGC 1860

AGGGATGGTT GATGCGATTT CCAAGTGTTC TGGATTTGTA AAGCGTTTGA TATAGTTGAC 1920

ATTGCTGTTA AAGAAGGCTT CCAAGTCGCT ATCGCTTGTC TTAACAAGGA TGGTGATAGG 1980

CTTGCTTGGT GCTACATTTA CTTCCGCACG CGCATTCCGA ACAGCACGAA TCAAGTCTTT 2040

GAGACTTTCC ACACCAGTGT GAGCCGCAAG GTCTTCAAAG GCTAGATTAA CAGTTGGGTA 2100

TGCAGCTGTC ACGATAGAAC CTTCTGAGAT TTGTCCAAAG ATTTCCTCTG TCACGAATGG 2160

CATGATTGGG TGAAGGAGAC GAAGGATCTT GTCCAGCGTA TAGAGGAGAA CAGATCGAGT 2220

AATGACCTTA TCGTCTTCAT TGTCGCTGTA TAGAACTTCC TTGGTCAACT CAACATACCA 2280

GTTGGCAAAT TCTTCCCAGA TGAAGTTGTA AAGGATATGA CCAGCCACAC CAAACTCGAA 2340

CTTATCAAAG TTTTCAGTAA CTTTTGCAAT GGTTTCGTTG AGATTGTGGA GAATCCAGCG 2400

GTCCGTCACA TTACCAGCCT CACCTGTTGC AACTTTTGTG ACATTGTCAT GCGCCACATC 2460

CAGCGTCAAA CCTTCATTGT TCATGAGGAT ATAGCGAGAA ATGTTCCAAA TTTTGTTAAT 2520

AAAGTTCCAT GAAGCATCCA TTTTCTCGTA AGAGAAACGA ACGTCTTGAC CTGGTGCGGA 2580

ACCGTTTGAA AGGAACCAAC GAAGGGCATC AGCACCGTAT TTCTCGATGA CATCCATTGG 2640

GTCAATCCCG TTACCGAGAG ATTTAGACAT CTTGCGTCCT TGCTCGTCAC GGATGAGACC 2700

GTGGATAAGC ACGTTTTGGA ATGGCTGACG ACCAGTAAAT TCCAAGGACT GGAAGATCAT 2760

ACGAGACACC CAGAAGAAGA TGATGTCGTA ACCTGTTACC AAGGTTGAAG TTGGGAAATA 2820

ACGTTTAAAG TCTTCTGAGT CGACTTCAGG CCAGCCCATG GTTGAAAATG GCCAGAGGGC 2880

AGAACTGAAC CAAGTATCCA AGACGTCTTC GTCCTGAGTC CATCCGTCAC CTTCTGGAGC 2940

TTCTTCGCCG ACATACATTT CACCATCAGC ATTGTACCAG GCAGGGATTT GGTGACCCCA 3000

CCAAAGCTGA CGAGAGATAA CCCAGTCGTG GACATTTTCC ATCCATTGAA GGAAGGTATC 3060

GTTGAAACGA GGTGGGTAGA ATTCGACCTT GTCCTCTGTG TCTTGGTTAG CAATGGCGTT 3120

CTTAGCCAAT TGGTCCATCT TGACGAACCA TTGAGTAGAC AAGCGTGGCT CAACTACGAC 3180

ACCTGTACGT TCTGAGTGAC CAACACTGTG GACACGTTTT TCGATTTTGA CAAGGGCACC 3240

GATTTCTTCC AACTTAGCAA CGACTGCCTT ACGAGCTTCA AAACGATCCA TGCCTGAAAA 3300

TTCAAAGGCA AGCTCATTCA TAGTTCCGTC GTCGTTCATG ACGTTGACTT GTGGCAAGTT 3360

ATGACGTTGG CCAACCAAGA AGTCATTTGG ATCGTGGGCA GGTGTGATTT TCACGACACC 3420

AGTACCAAGC TCAGGATCTG CGTGCTCATC TCCAACGATT GGGATGAGTT TATTAGCGAT 3480

TGGAAGGATG ACGTTTTTAC CAATCAAGTC CTTGTAGCGC GGGTCTTCTG GATTAACCGC 3540

AACCGCAACG TCCCCAAACA TAGTCTCAGG ACGAGTTGTA GCAACTTCAA GGGCGCGTGA 3600

ACCATCTTCC AGCATGTAAT TCATGTGGTA GAAGGCACCT TCTACATCCT TGTGAATCAC 3660

CTCAATATCA GAAAGGGCTG TGCGAGCTGC TGGGTCCCAG TTGATTATAA ACTCACCACG 3720

ATAGATCCAG CCTTTCTTGT AAAGGTTCAC AAAGACCTTA CGAACAGCTT TTGACAAACC 3780

TTCATCAAGA GTGAAACGCT CACGAGAATA GTCTACAGAA AGCCCCATCT TGCCCCATTG 3840

TTCCTTGATG GTAGTGGCAT ATTCGTCTTT CCATTCCCAG ACCTTCGTCA AGAAAGACTC 3900

ACGACCTAGG TCATAACGCG TAATACCCTC ACCACGTAAG CGCTCCTCAA CCTTAGCCTG 3960

AGTCGCAATA CCAGCGTGGT CCATACCTGG AAGCCAAAGG GTATCAAAGC CTTGCATGCG 4020

TTTTTGACGG ATGATGATAT CCTGCAAAGT CGTATCCCAA GCGTGACCAA GGTGAAGTTT 4080

CCCAGTTACG TTTGGTGGTG GAATCACGAT TGAATAAGGC TTAGCCTTTT GATCGCCTGA 4140

AGGCTTGAAA ACATCCGCAT CAAGCCATTT TTGGTAACGA CCAGCCTCAA CCTCGGCTGG 4200

ATTGTATTTA GGTGAAAGTT CTTTAGACAT GTGTGTGTCC TTTCTCTATT TTGTTTATTT 4260

TATTTTGAAT TTGCTTAGCA GCTTCTTCTG CAGACAAATT CGTATTATTT ATTTTAAAGT 4320

AGTGGTGCAA CTCATTCGGT TGATGTTGGG AATTTAATTG AAGTGTTTCA GCGGTCTCTA 4380

AAATTTCTCT TTCAGATACC TCAATATGTC GTTTTAAGGG TTTGTGCTTT AATCGATTCT 4440

CCGTTCGATT TCGACGTATG CACTCTTCAA GACTTGTTTC CAATTCAACA AACAGAATCT 4500

CTTGATGAAA GTTATCCAAT AAATCCTGAA TTTGCTTTAA ATACATCAGC TGGTACTGAT 4560

TTGAAAAATC AATTACGTCT GTTAAAATTA CTGATCGCTG ATTTCTTGCA CTTGCTCCAA 4620

GGAAAGAAAA GGTAATTCCA CGAACAAATT CCCACATCTC CTCGGTATAA TCCTGATAGA 4680

TCTCTAGTGC AAAATCAATG GCTTGATGGT TATAAAATAG GGTAGCATCC GTCAGTCGAG 4740

ATAATTCTTG ACCAATGGTC ATTTTTCCTG ATGCTGGAGC ACCAATGATG AAAAGATGCA 4800

TCAAATCACC TCCCACTCAC TCCTCAGCAA GCCATATCTC AAATCATCAC AGCAGTTGCC 4860

TTGAGCATCT TTGCGGTCTC TTATGCGAGC TTCGAGGGTA AAGCCAAGCT TTTCCGAGAC 4920

TCGTTGACTT TGAAGGTTAT ATCCAAAGCA AGTTAGTTCA ATCTTGTGAA GACCAAGTTC 4980

TTTAAAAGCT AGATCAATCA AGGAACACGC TGCTTCTGGA ACATAACCTC GACCCCAATA 5040

GTCTGGGTGC AAGGTATAGC CAAGCTCTAG CACATCATCC GCATGAAGAT GGTTGAAGTC 5100

AACAGAACCA ATGACTTTAT CGGTTCCTTT GACGACAATC CCATAGCCAG CTGGGAGATT 5160

TTCCTTTTGA GTACGCTCCG GAAGAATGTG CTCCAGATAA TAAATCTCAT CTTCCAAGAT 5220

CTTGACTGGA GGAAAACCTG CTGGATAGGC GACCTCTGGC AAACTAGCGT AGGTATGGAT 5280

ATCCTCAGCA TCCACCACTG TGCGGACTCG TAAAACGAGA CGTTCTGTTT CGATTTTATC 5340

TGGCAGCTCA GTTCTTGCCA TCCTTCTTCC TCGCTTTTTT GATGAAACTG CCCTTCATAT 5400

CTACACGCTT GTCCAGATAG CGATAAACGC GCTGATATCC ATCTCCCATG AAATAGGTTG 5460

GGGCAAACAG TTGATTTTTA AAATGTCCCT TTTCATCCAG GAGTTCTGGG GCAACAAGTC 5520

GCTCAAGAAT CTTGGCAAAG ATGTGGCAAA TACCGTCTTC CTCAACAATC CTATCTACCC 5580

GACAATCTAA AACAAGTGGA CAGGCGTCTA AAATAGGAGT CTGAGTTCGT TCAGAAATTT 5640

CATAATGCAC TCCCAAACGT TCCAATTTCT CCTGATGACT GATAAAACCA GCCTGCTCCA 5700

TCGCAAGCAT AGAAGTTTCA TCAGAAATAT TCACAGTAAA TTTTTGATAC TGTTTGATCT 5760

GCTCTGCGGC ATTCTCTCTC GCAACGACTC CAATCACAAC CCAATCTCCT AGACTATAAG 5820

AGGAACTACA GGTCGTGATG TTATAGCCAA AATTCTAATC TTGATATCCT AAAATAAAAA 5880

CAGGAAAACC ATAATATAGT TTACTTGTGT TAAAAGATTG CTTCATAACA ACCCCCTTTG 5940

ACTAAGACGT AAAAGAAAAG CCCTGCCATC TACATGACAG GGACGAATGT GTTTATCCGC 6000

GGGG 6004

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5857 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

TGTAGAATTC ACGACAATGC TTCGTTGATT TCTGGGTTGA TTTCGTCGCG TTCTGGCAAG 60

CGAGTCAATG AACCAAAAAT AGTACACAAT GTGGTATAAT CCTTTTATGG CATATTCAAT 120

AGATTTTCGT AAAAAAGTTC TCTCTTATTG TGAGCGAACA GGTAGTATAA CAGAAGCATC 180

ACACGTTTTC CAAATCTCAC GTAATACCAT TTATGGCTGG TTAAAGCTAA AAGAGAAAAC 240

AGGAGAGCTA AACCACCAAG TAAAAGGAAC AAAACCAAGA AAAGTTGATA GAGATAGACT 300

TAAAAACTAT CTTACTGACA ATCCAGATGC TTATTTGACT GAAATAGCTT CTGACTTTGG 360

CTGTCATCCA ACTACCATCC ACTATGCGCT CAAAGCTATG GGCTACACTC GAAAAAAGAA 420

CCACACCTAC TATGAACAAG ACCCAGAAAA AGTAGCCTTA TTTCTTAAGA ATTTTAATAG 480

TTTAAAGCAC CTAACACCTG TTTAGATTGA CGAAACAGGA TTCGATACTT ATTTTTATCG 540

AGAATATGGT CGCTCATTAA AAGGTCAGTT AATAAGAGGC AAAGTATCTG GAAGAAGATA 600

TCAGAGGATT TCTTTGGTTG CAGGTCTAAC AAATGGTGAG TTAATCGCTC CAATGACTTA 660

CGAAGAGACG ATGACGAGCG ACTTTTTTGA AGCTTGGTTT CAGAAGTTTC TCTTACCAAC 720

ATTAACCACA CCATCGGTTA TTATTATGGA TAATGCAAGA TTCCATAGAA TGGGGAAGCT 780

AGAACTCTTG TGTGAAGAGT TTGGGTATAA ACTTTTACCT CTTCCTCCCT ACTCACCTGA 840

GTACAATCCT ATTGAGAAAA CATGGGCTCA TATCAAAAAG CACCTCAAAA AGGTATTACC 900

AAGTTGCAAT ACCTTTTATG AGGCTTTTTT GTCTTGTTCT TGTTTCAATT GACTATATAA 960

ATTGTCTAAG CGAAACAACC GATAAGAATT GGCACAAAAG CGACCGTATT TTTGTTACCA 1020

ATACAGGAAA AACAGTTCAT AGTTCTATCT TGAGCAAGTC TCTCCAGCGA GCAAACGAAC 1080

GCCTTAAAAA ACCAATTCCC AAACATCTGT CCCCTCACAT CTTCAGACAC ACCACTATTA 1140

GCATCTTATC AGAAAATAAA ATTCCTTTAA AAACAATCAC GGACAGGGTT GGTCATCCCG 1200

ACTCTGAAGT CACTACTTCC ATCTACACCC ACGTCACAAA GAACATGAAA GATGAAGCAA 1260

TCAATGTACT GGATAAAGTT ATGAAAAAGA TTTTTTAAAA AGTTTTGTCC CTTTTTTGCC 1320

CTCTAAATAC AAAAATAGCC CTTCGGATAA AATCCGAGGG GCTAGAAACG TTGTTAAATC 1380

AACGGCCGAA CTTTTGAATT TCATGGTTCG GGATAAAATA GTTCACTGAA CTATTTTATT 1440

TTTTAAGGTT ATCATAATAT CAAATAGTTC AATTAAATAC GCTAAATTAC TAATATACTT 1500

TTTACCTTTT TCATTCTAAA ATGTAAAGTA CAAACAATTA CAATATACTA GAGGGGGAGT 1560

AAAAAAGGTA TTAAATCGAT GAGTTCAGCA GGCAAGAAAA TAGCACCTTT ACGGGTGCTA 1620

TTTTTTAATT AACGCCACGT TAACTTTTGA TTGATGAATT TTATTGTTTG GCACTTCTTT 1680

CATTTCACGG TAAACATCGA TGAAATTCTT TCCAACATTA TTTTTGGAGT TAACTGCATT 1740

TATTTTTGTA TTAATAACTT TTTTAGTATC GAAAGAATGG TTTAAGAAAT CCATAACTAA 1800

CTCTCCTTTC TCATCCTGTA ATCAAGATTT TTATCAATGT CAAAATAGTA TTTTCTATCA 1860

ATCCAAATTG GTCCTTCTCC TTTAGAAATA GCAAGTACAT CTACCGGACC TCCTACTGTT 1920

TCAAGAGTGT TGACAATTTT TCTCTTAAAT GAAGTTAATT CAATAAATGT TTTAGCTGTA 1980

CTCGCCATTT CATTAAGTGG TTGCATTCCA ATAAGGTCTA TTATAGGATT TATATAATAT 2040

TTTTGCTGTA TAGATGATAT ATTTTCAAAT ATATTCTCAA TTTCATCACC CAATCCATTT 2100

TTCTCCATAA CTGATGATAC TTGCTCTGCG ATATATACAT TTAAGTTAGG ATCTATACCA 2160

TTCATAATCG TCTCAACCAT CTCTGACTGT GCAAAAGGGA TTATATGACA AGTTTTATGA 2220

TGATTTATCA CACTTTCATT AATAACTTTC CAAATTAATC GTTTAGAAAA AATTCCATAT 2280

AATTCAATTT GTCTTATAGA TGGAAATATC TCGTCTGTAC CATAACCTGC TATAACTAAT 2340

CCAGTTATGT TTGTTGAGTC ATATCCAATG AAAATCGCTT TATATAAAGA TTTAGCAATA 2400

ACTTCAACCT CATCATCAGT ATGAGGAAAG GATTTAAAAA CATCGTCTAC AATGCTTTTT 2460

ATTAACTCTA ACTCAGCTTC AAAAAATTCA AAATTACTTT CAGCTTCTAC TTTTGAAATT 2520

TCTAAACTAA AATTAGTTAT AGCATTTAAT AAAATTTTAT TAAAATCATC TAGAGTGATG 2580

GTTTCACCAT TAGAAACTCT TAAATCAGCT GTTTCTTGCG CTTCATAGGC AATGCTGTCC 2640

AAAATACTTC TTGTACTTCT GACAATATAA TTTCTTAATA AATCCTCAAC TTGTAGATGT 2700

TTAAAGGAAA TTAAAAATTC TATTAGCTTT TCAACGTATT GGGCAGTATT ATCTAATAAA 2760

TCTGTGCCAA TAGCCTGCTT AAACTCATTT AAAATTACCT CCCACGGAAT TTCCATAAAC 2820

GAAGCGTTCC CATATATCAT GATCCCCACG GAATGTTCTT TTGATAAAGT GAATAATTTT 2880

CGGGCGCTAT TAAAAACTTT TGAATTTTTC CCGTCTGATA AGGTTACAGC GCTATCAGAA 2940

GCCAATACAA CACCATTTTT ATTTAATATT CCAATTTCTG CTGTCAAAAT ATCACCTAAA 3000

CTTTCTAAAC CTGCTCATGC TCTAATGGTA CAACAGCTAA GGTCTTACCA AGACTTGCCA 3060

ACACTTTTAA TACTGTATCA AGTTGTGGGC TTGTCTTTCC TGTTTCCATT CTAGCGATAA 3120

CTGGCTGACT AACACCGCTC ATCTCCTCTA GTTTCTTCTG ACTAATACCC TTTTCATTTC 3180

TAGCCTCGAT AAGCTCACTC ATGATAGCCA CGCGCATATC ACTTTCCAAA ATTTCCTCTT 3240

TGCTGAATAA TTCAGCTCTT ACATCTTTCC AGTTACTACC AATAGCATTA TTTTTCATTG 3300

TCTAAACCTC TTTCTTTTAA ATCTGCAAGT TCACGTTTAG CTTGCTCAAT CTCTCTTTTG 3360

GGTGTTTTCT GTGTCCTTTT CATAAAATGA TGCAGTAAAA CAAAACTACC ATCCATCCAA 3420

GCAACAAATA AAATTCTATC TCTAAGTGGT CTCAGCTCCC AAATTTCAGC ATCTAAATGC 3480

TTAATATATG GTTCGCCTGC GCGTGTTCCA TGTTGGCTTA ACAACTCAAT ATAATCATTA 3540

ATTTTATTAA GCTTAATTCT GCTATCTTTC CCTTTTTTAC TGGTAAGCTC TCGCATATAA 3600

TCAAAAACAG GCTCATTGCC GTTTTTATCC TTGTAAAAAT AGATATTATG CACTATTAAC 3660

ACCTCTTCCT AATAACAATT ATAACCTAAA AGTTATTGTT TGTAAATACT TTTAAGTTAT 3720

TAAAATAAAA AGCACCTAGT TTCCTAGATG CTAGCACAAT GACACGGATT CGCACCGTGG 3780

CTACCTCTAT CAAGGTGTAC TCCTTCTATA CTATCCCTTG TGCTTTAGAA TATTATACCA 3840

CACAATCAAC TAGATACCTA CCATCTCATG ATATACCCCC ATTTTGGGCA AGGGTACAAC 3900

GCTAAAATAC AAATCAGAAT AGATATTAAA CCACTTATTT AACTTATCAT AAGCTGGTGA 3960

TTGACTGATA AATAATATCC GCTGACAAGC TCCGATAACA TTCATGTGAT TGTACACATA 4020

AACCTCTTTT ACAGCCTCTA AAATGTCAGC CTCACTTGTT TGTACCCTAA TATCTGTTAT 4080

CTGCTTGATA GTTGCGTATT TTTGATAAGC TAGCATATCT TGATTTTTAG CAGCATCAAA 4140

CATTTTACGC TCAAGGACAC TATACTTAGG TTGTTCTTTA TCTCGCATGA AATACCACTT 4200

GAGCCATAAA ATCTTTTCTC GGTGTATTAC AGAAATACGC TCAATTTTCT TCTTTGTCAT 4260

TGCTACCTCC TAAATCATCA ATTTAACAAT TCTAACCACT CACTTTTAGA AATAGTTGCA 4320

TAGATCTTGT TCGATGTATG ATACAAAGGT TCTAAATCTT TTTCCACCCT AATATAGTTC 4380

ATCTTATCCT CATGAGTAGG AAAGTATAGT ATTTCCGTTT CATCCTCGTT TAGGATACGA 4440

TTGCACCAAT CATCAATAAT AACTGGCACT TCCCACTCAC GCCATTTTTT AAGGTTTTCT 4500

AAAAGTTCAT TATCACTAAA TAGCTCGCCA TCTATTTGGA AAAATTCCCC TAAGTCATTG 4560

TTTCCTTCAA CAATAATAAA CTCTGGCATA TTTCTATTAC TTAATAACTC CTTGAGTTCT 4620

TGTAACTCTT TGATTTCCTT TAGATACTTC CTCAATTTCC AACCTCAATT CTTCAATCTG 4680

CCTTACTACT CCAAAAATTT CATGGGTCTT ATAAGATTGT TCAAGTATAG CCTTTGCTGC 4740

TTGAGTTCTT ATAAACGGGT TGACCTTACT GTCCATCATA ATATCATTGA GTACAGAAAC 4800

AGCGTTAGAT GATGCTAAAT AAAGCATTTG AGTTGTTTTA TCCATCATCT CATCTTGCTT 4860

TATCCTCAAT GTCTTTTTAA CCGCTGCAAC TTTTAGATAC TTATGACCTG TTGCGCGTGA 4920

TACCCCTGCT TTTTGACATG CTTTGTCTAT CGTTGGCTCG GTAAGCATGG CATCTATGAA 4980

TTTAATTTGC TTGGACGTAA GGTTATCATT TTCATTTCCT GCCATCTATT ACCTCCTCAT 5040

TATCAAAATA AAGGGTTGCC CCTTTATTTC CCTATGCTAG ATAATTCTGC AATTCTGCAT 5100

CCATTGCCTC TGAATTGCCC TCAACAATCA TTTCATGCTG TACTAAATCA ATCTTATCTC 5160

CGTTAATAAG TAAACCACCG TGGAAATAAT CAATTTTTCT ATCAAGGAAA TGTACTAGCT 5220

TTTCAAGGCG TTGCTGTTGG CTGAATTGCT CCATGTCAAT TTCGATATAA GCAAGGGTAG 5280

TATCATTATC CATAATATCT TCTAATTTTC TAAGAGCTAG AGGTTTATTT TTATATTTTT 5340

CTAGGTATTC TCTCATTTCT GCCACTGTTA ATTTGATACT AGATAATAAA CTTAGTTCAG 5400

CTGCATCATC TGCTGTAATA GGCTCTTCTT TTGATTCATG GTTTGCTAGT TCAGCATTTT 5460

TCTCTTTTTC TAGTTGCTGA TACAATAGCT GAGCAGTATT TTGGGAATAG TTTTCGCCCT 5520

CTTTTTTATA TTTTAAAAGT TCTTGCTCTG CATACACTTT CCCGATAATC ACTTCCTTAT 5580

AAACTAATTG CCCATCTTGA GCTTTTAGCT TAATACTCCC ATGCTCTGGA ATTTCAATAT 5640

ACTTAATTAT ACCATTTTTT GAGTATAAAA CAAAGCCTTT CTCCATCATT TTTAATAATT 5700

TATCATCCTT GTTTTCAGTC ATGCTTTTCT CCTTTATTTC ATTTTATTAT AATCTGAATA 5760

CCCCTAGTCT ATTTATTTCA CTAGGTTTTT AGGGTTCGTA TGCTAAAATA CTACCCTTTT 5820

TGTGTACCTT ATGGCTGACT TTTCAAATTG GTTAGTT 5857

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10254 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

AAAATGATAG CAGGAGAGTT TTCCCGTCCA TCAGACCCAG AACTGAGAGC CTTAGCTCAG 60

GCTTCTCGCC AAAAACAGGC CGCCTTTAAC AAGGAAGAGA ACCCCTTGAA GGGAGCCGAA 120

ATCATCAAGA CTTGGTTTGC CTCAACCGGG AAAAATCTTT ACATCAACAC TCGCTTGATG 180

GTGGACTACG GTGTCAACAT CCATCTAGGG GAAAATTTTT ATTCTAATTG GAACTTGACC 240

ATGCTGGATA TCTGTCCCAT TCGTATCGGG GACAATGCTA TGATTGGTCC TAATTGTCAG 300

TTTTTGACAC CCCTCCATCC ACTAGATCCA CAGGAACGCA ATTCAGGTAT CGAGTACGGA 360

AAGCCTATCA CAATCGGAGA TAATTTCTGG ACTGGTGGTG GCGTCATTGT CCTTCCTGGA 420

GTGACACTGG GAAATAATGT CGTTGCAGGA GCAGGGGCAG TAATTACCAA ATCTTTTGGC 480

GACAACGTTG TCCTAGCTGG CAATCCTGCG CGCGTGATTA AGGAAATACC TGTTAAATAG 540

AAGTAAAAAG GAACAGCTGG GGTTGTTTCT TTTTTGTAGG TTTCATCATT TTTTACCCAG 600

TTCACATTTA CCTACTCTAT CTCTTAGCAA GTCTGTTTCA TTAAGCAAGT TCAAAGCATC 660

TCGTAAGTGG GATGTTTTTC TCCTCAGTTC ATCAGCTTCC TCCTTGACAC TCGGTCAGAT 720

TTTGATACAA TAGTACAAAA TTAGAGGAGG CAGGCTATGA TTCAGAAACA TGCGATTCCT 780

ATTTTAGAGT TTGATGACAA TCCTCAGGCG GTTATCATGC CCAATCACGA GGGGCTGGAC 840

TTGCAGTTGC CAAAGAAGTG TGTTTATGCA TTTTTAGGTG AGGAGATTGA CCGCTATGCG 900

AGGGAAGTAG GGGCGAACTG TGTTGGCGAA TTTGTTTCTG CCACCAAGAC CTATCCAGTT 960

TATGTCGTGA ACTACAAGGA CGAGGAGGTC TGTCTGGCTC AGGCTCCTGT TGGCTCCGCT 1020

CCAGCAGCCC AGTTTATGGA TTGGTTGATT GGCTATGGTG TGGAGCAGAT TATCTCTACT 1080

GGGACCTGTG GTGTCCTAGC TGATATAGAG GAAAATGCCT TTCTAGTCCC TGTTCGCGCT 1140

CTGCGAGATG AAGGAGCCAG TTACCACTAT GTGGCACCTT GTCGTTATAT GGAAATGCAG 1200

CCAGAGGCTA TTGCTGCTAT TGAGGAAGTT TTGGAAGACA GAGGGATTCC TTATGAAGAA 1260

GTCATGACCT GGACGACAGA CGGTTTTTAC CGAGAAACGG CTGAAAAGGT GGCTTATCGT 1320

AAGGAAGAAG GCTGTGCTGT TGTGGAGATG GAGTGTTCTG CTCTTGCGGC AGTAGCTCAA 1380

TTGCGTGGGG TTCTCTGGGG TGAATTGTTG TTCACAGCAG ATTCTCTAGC GGACTTGGAC 1440

CAGTACGACA GTCGTGACTG GGGCTCGGAA GCTTTTAATA AGGCGCTAGA ACTGAGTTTA 1500

GCAAGTGTTC ACCACCTTTA GTTGTACTGG CAAAGGATTT GTTTTATCAT AAAATGTCTA 1560

GCTCATACTT TTCAAAAATA TGTTTAAACG AGGTCACCTT CCTCTTGTCC TAGGCATGTT 1620

GAGGTTGGGA AAAATCTTTA AAATCAGAAA AACGTATCAT ATCAGGTGAT GAAAACTTTG 1680

ACACTATGCG TTTTATGTCG ATAAGATTTA GAGTGAGATG AAATGATACT CTTCGAAAAT 1740

CTCTTCAAAC CAGGTCAGCT TCACCTTGCC GTAGGTATAT GTTACTGACT TCGTCAGTCT 1800

TATCCGGCAA CCTCAAAACG GTGTTTTGAG CTGACTTCGT CAGTTCTATT TGCAACCTCA 1860

AAACAGTGTT TTGAGCAACC TGTGACTAGC TTTCTAATCG ATGCCTTGGT TTTCATTGCC 1920

TATAATCAAA AAGAGAAATT TTCTCCTGAA AAGCATATAG AGTAGCTGGC GTTAAAAGCT 1980

CCTGTCTTGC TTTTTTGACC TATAGTCACA TCTATCAAGT ATTGTTCTTG CCTAAGCTAT 2040

CAATAAAAAG GTGGCATTTT TTAGGCTTGG TGTTAGTAGA TTTTGCCTTA TCCTATCTAA 2100

GTCATTTCGA ACTTTTTATG GTACAATGGA AACATGTTAT TCAAATTATC TAAGGAAAAA 2160

ATAGAGCTAG GCTTATCTCG TTTATCGCCA GCCCGTCGTA TTTTTTTGAG TTTTGCCTTG 2220

GTCATTTTAC TAGGCTCTCT TCTTTTGAGC TTGCCCTTTG TCCAAGTTGA AAGCTCACGA 2280

GCGACTTATT TTGATCATCT TTTCACTGCT GTCTCTGCAG TCTGTGTGAC GGGTCTCTCA 2340

ACCCTTCCAG TAGCTCACAC CTATAATATC TGGGGTCAAA TAATCTGTTT GCTCTTGATT 2400

CAGATCGGTG GTCTAGGGCT CATGACCTTT ATTGGGGTTT TCTATATCCA GAGCAAGCAA 2460

AAGCTTAGTC TTCGTAGCCG TGCAACTATT CAGGATAGTT TTAGTTATGG AGAAACTCGA 2520

TCTTTGAGAA AGTTTGTCTA TTCTATTTTT CTCACGACCT TTTTGGTTGA GAGCTTGGGA 2580

GCTATTTTGC TTAGTTTTCG CCTTATTCCT CAACTTGGCT GGGGACGTGG TCTTTTTAGT 2640

TCCATTTTTC TAGCGATCTC AGCCTTCTGT AATGCCGGTT TTGATAATTT AGGGAGCACC 2700

AGTTTATTTG CTTTTCAGAC CGATTTACTG GTCAATCTGG TGATTGCAGG CTTGATTATT 2760

ACAGGCGGCC TTGGTTTTAT GGTCTGGTTT GATTTGGCTG GTCATGTAGG AAGAAAGAAA 2820

AAAGGACGTC TGCACTTTCA TACGAAGCTT GTACTATTAT TGACTATAGG TTTGTTGTTA 2880

TTTGGAACAG CAACTACTCT CTTTCTTGAG TGGAACAATG CTGGAACGAT TGGCAATCTC 2940

CCTGTTGCCG ATAAGGTTTT AGTTAGCTTT TTTCAAACAG TGACGATGCG AACAGCTGGC 3000

TTTTCTACGA TAGATTATAC TCAGGCTCAT CCTGTGACTC TTTTGATTTA TATCTTACAG 3060

ATGTTTCTAG GTGGGGCACC TGGAGGAACA GCTGGGGGAC TCAAGATTAC GACATTTTTT 3120

GTCCTCTTGG TCTTTGCACG AAGTGAGCTT CTAGGCTTGC CTCATGCCAA TGTTGCGAGA 3180

CGAACGATCG CGCCGCGAAC GGTTCAAAAA TCCTTTAGTG TCTTTATTAT CTTTTTGATG 3240

AGCTTCTTGA TAGGATTGAT TCTGCTAGGG ATAACAGCCA AAGGCAATCC TCCCTTTATC 3300

CACCTCGTAT TTGAAACCAT TTCAGCTCTT AGTACAGTTG GTGTAACGGC AAATCTGACT 3360

CCTGACCTTG GGAAATTGGC TCTCAGTGTT ATCATGCCAC TTATGTTTAT GGGACGAATT 3420

GGTCCCTTGA CCTTGTTTGT TAGCTTGGCA GATTACCATC CAGAAAAGAA AGATATGATT 3480

CACTATATGA AAGCAGATAT TAGTATTGGT TAAGAAAGGA AAGAGCATGT CAGATCGTAC 3540

GATTGGAATT TTGGGCTTGG GAATTTTTGG GAGCAGTGTC CTAGCTGCCC TAGCCAAGCA 3600

GGATATGAAT ATTATCGCTA TTGATGACCA CGCAGAGCGC ATCAATCAGT TTGAGCCAGT 3660

TTTGGCGCGT GGAGTGATTG GTGACATCAC AGATGAAGAA TTATTGAGAT CAGCAGGGAT 3720

TGATACCTGC GATACCGTTG TAGTCGCGAC AGGTGAAAAT CTGGAGTCGA GTGTGCTTGC 3780

GGTTATGCAC TGTAAGAGTT TGGGGGTACC GACTGTTATT GCTAAGGTCA AAAGTCAGAC 3840

CGCTAAGAAA GTGCTAGAAA AGATTGGAGC TGACTCGGTT ATCTCGCCAG AGTATGAAAT 3900

GGGGCAGTCT CTAGCACAGA CCATTCTTTT CCATAATAGT GTTGATGTCT TTCAGTTGGA 3960

TAAAAATGTG TCTATCGTGG AGATGAAAAT TCCTCAGTCT TGGGCAGGTC AAAGTCTGAG 4020

TAAATTAGAC CTCCGTGGCA AATACAATCT GAATATTTTG GGTTTCCGAG AGCAGGAAAA 4080

TTCCCCATTG GATGTTGAAT TTGGACCAGA TGACCTCTTG AAAGCAGATA CCTATATTTT 4140

GGCAGTCATC AACAACCAGT ATTTGGATAC CCTAGTAGCA TTGAATTCGT AAAGAGGGAT 4200

GACCCCTCTT TTTTGATGCC TAAGATGGCA AATAGAGACA GAAGCCCCTT GTCTTCTAGT 4260

AAAAGTTCTT CAAAGGCTGG ACTTTATGGT AAAATAGAAA GAAGTGACAA GAGAGAGTAA 4320

TACTCAATGA AAATCAAAGA TCAAACTAGG AAACTAGCTA CGGGCTGCTC AAAACACTGT 4380

TTTGAGGTTG CAGATAGAAC TGACGAAGTC AGTAACATCT ATACGGCAAG GCGACGTTGA 4440

CGCGGTTTGA AGAGATTTTC GAAGAGTATA AGAAAAAATC AGTCCCCTAA AGGAGTAGAT 4500

TATGAAGTTA TTGTCTATCG CAATTTCTAG CTATAATGCA GCAGCCTATC TTCATTACTG 4560

TGTGGAGTCG CTAGTGATTG GTGGTGAGCA AGTTGGGATT TTGATTATCA ATGACGGGTC 4620

TCAGGATCAG ACTCAGGAAA TCGCTGAGTG TTTAGCTAGC AAGTATCCTA ATATCGTTAG 4680

AGCCATCTAT CAGGAAAATA AATGCCATGG CGGTGCGGTC AATCGTGGCT TGGTAGAGGC 4740

TTCTGGGCGC TATTTTAAAG TAGTTGACAG TGATGACTGG GTGGATCCTC GTGCCTACTT 4800

GAAAATTCTT GAAACCTTGC AGGAACTTGA GAGCAAAGGT CAAGAGGTGG ATGTCTTTGT 4860

GACCAATTTT GTCTATGAAA AGGAAGGGCA GTCTCGTAAG AAGAGTATGA GTTACGATTC 4920

AGTCTTGCCT GTTCGGCAGA TTTTTGGCTG GGACCAGGTC GGAAATTTCT CCAAAGGCCA 4980

GTATACCATG ATGCACTCGC TGATTTATCG GACAGATTTG TTGCGTGCTA GCCAGTTCTA 5040

ACTGCCTGAA CATACTTTTT ATGTCGATAA TCTCTTTGTC TTTACGCCCC TTCAGCAGGT 5100

CAAGACCATG TACTATCTGC CTGTCGATTT CTATCGTTAT TTGATTGGGC GTGAGGACCA 5160

GTCTGTCAAT GAGCAAGTGA TGATTAAGTG CATTGACCAG CAACTCAAGG TCAATCGACT 5220

CTTGATAGAC CAACTTGATT TGTCCCAAGT GAGTCATCCC AAAATGCGAG AATATCTGCT 5280

GAATCATATT GAACTCACGA CGGTGATTTC CAGTACCCTG CTCAACCGAT CTGGAACAGC 5340

GGAGCATCTG GCAAAAAAAC GCCAATTGTG GACCTATATT CAGCAGAAAA ATCCAGAAGT 5400

CTTTCAGGCT ATTCGTAAGA CCATGTTGAG CCGTTTGACC AAACATTCTG TCTTGCCAGA 5460

TCGCAAACTG TCCAATGTCG TCTATCAAAT CACCAAATCT GTTTATGGAT TTAATTAATA 5520

TAAGTGTTTT ATAAGAGGGA TTTAAGAAAA ATTTTAACTT TTTCTTAGTC CTTTTTAATT 5580

TCAGGAGATT ATACTAGAGT CATCAAATAA AGAAAGACTC TAAGGAGAAT CCTATGAAAT 5640

TCAATCCAAA TCAAAGATAT ACTCGTTGGT CTATTCGCCG TCTCAGTGTC GGTGTTGCCT 5700

CAGTTGTTGT GGCTAGTGGC TTCTTTGTCC TAGTTGGTCA GCCAAGTTCT GTACGTGCCG 5760

ATGGGCTCAA TCCAACCCCA GGTCAAGTCT TACCTGAAGA GACATCGGGA ACGAAAGAGG 5820

GTGACTTATC AGAAAAACCA GGAGACACCG TTCTCACTCA AGCGAAACCT GAGGGCGTTA 5880

CTGGAAATAC GAATTCACTT CCGACACCTA CAGAAAGAAC TGAAGTGAGC GAGGAAACAA 5940

GCCCTTCTAG TCTGGATACA CTTTTTGAAA AAGATGAAGA AGCTCAAAAA AATCCAGAGC 6000

TAACAGATGT CTTAAAAGAA ACTGTAGATA CAGCTGATGT GGATGGGACA CAAGCAAGTC 6060

CAGCAGAAAC TACTCCTGAA CAAGTAAAAG GTGGAGTGAA AGAAAATACA AAAGACAGCA 6120

TCGATGTTCC TGCTGCTTAT CTTGAAAAAG CTGAAGGGAA AGGTCCTTTC ACTGCCGGTG 6180

TAAACCAAGT AATTCCTTAT GAACTATTCG CTGGTGATGG TATGTTAACT CGTCTATTAC 6240

TAAAAGCTTC GGATAATGCT CCTTGGTCTG ACAATGGTAC TGCTAAAAAT CCTGCTTTAC 6300

CTCCTCTTGA AGGATTAACA AAAGGGAAAT ACTTCTATGA AGTAGACTTA AATGGCAATA 6360

CTGTTGGTAA ACAAGGTCAA GCTTTAATTG ATCAACTTCG CGCTAATGGT ACTCAAACTT 6420

ATAAAGCTAC TGTTAAAGTT TACGGAAATA AAGACGGTAA AGCTGACTTG ACTAATCTAG 6480

TTGCTACTAA AAATGTAGAC ATCAACATCA ATGGATTAGT TGCTAAAGAA ACAGTTCAAA 6540

AAGCCGTTGC AGACAACGTT AAAGACAGTA TCGATGTTCC AGCAGCCTAC CTAGAAAAAG 6600

CCAAGGGTGA AGGTCCATTC ACAGCAGGTG TCAACCATGT GATTCCATAC GAACTCTTCG 6660

CAGGTGATGG CATGTTGACT CGTCTCTTGC TCAAGGCATC TGACAAGGCA CCATGGTCAG 6720

ATAACGGCGA CGCTAAAAAC CCAGCCCTAT CTCCACTAGG CGAAAACGTG AAGACCAAAG 6780

GTCAATACTT CTATCAAGTA GCCTTGGACG GAAATGTAGC TGGCAAAGAA AAACAAGCGC 6840

TCATTGACCA GTTCCGAGCA AAyGGTACTC AAACTTACAG CGCTACAGTC AATGTCTATG 6900

GTAACAAAGA CGGTAAACCA GACTTGGACA ACATCGTAGC AACTAAAAAA GTCACTATTA 6960

ACATAAACGG TTTAATTTCT AAAGAAACAG TTCAAAAAGC CGTTGCAGAC AACGTTAAAG 7020

ACAGTATCGA TGTTCCAGCA GCCTACCTAG AAAAAGCCAA GGGTGAAGGT CCATTCACAG 7080

CAGGTGTCAA CCATGTGATT CCATACGAAC TCTTCGCAGG TGATGGTATG TTGACTCGTC 7140

TCTTGCTCAA GGCATCTGAC AAGGCACCAT GGTCAGATAA CGGTGACGCT AAAAACCCAG 7200

CCCTATCTCC ACTAGGTGAA AACGTGAAGA CCAAAGGTCA ATACTTCTAT CAATTAGCCT 7260

TGGACGGAAA TGTAGCTGGC AAAGAAAAAC AAGCGCTCAT TGACCAGTTC CGAGCAAACG 7320

GTACTCAAAC TTACAGCGCT ACAGTCAATG TCTATGGTAA CAAAGACGGT AAACCAGACT 7380

TGGACAACAT CGTAGCAACT AAAAAAGTCA CTATTAACAT AAACGGTTTA ATTTCTAAAG 7440

AAACAGTTCA AAAAGCCGTT GCAGACAACG TTAAGGACAG TATCGATGTT CCAGCAGCCT 7500

ACCTAGAAAA GGCCAAGGGT GAAGGTCCAT TCACAGCAGG TGTCAACCAT GTGATTCCAT 7560

ACGAACTCTT CGCAGGTGAT GGCATGTTGA CTCGTCTCTT GCTCAAGGCA TCTGACAAGG 7620

CACCATGGTC AGATAACGGC GACGCTAAAA ACCCAGCTCT ATCTCCACTA GGTGAAAACG 7680

TGAAGACCAA AGGTCAATAC TTCTATCAAG TAGCCTTGGA CGGAAATGTA GCTGGCAAAG 7740

AAAAACAAGC GCTCATTGAC CAGTTCCGAG CAAACGGTAC TCAAACTTAC AGCGCTACAG 7800

TCAATGTCTA TGGTAACAAA GACGGTAAAC CAGACTTGGA CAACATCGTA GCAACTAAAA 7860

AAGTCACTAT TAAGATAAAT GTTAAAGAAA CATCAGACAC AGCAAATGGT TCATTATCAC 7920

CTTCTAACTC TGGTTCTGGC GTGACTCCGA TGAATCACAA TCATGCTACA GGTACTACAG 7980

ATAGCATGCC TGCTGACACC ATGACAAGTT CTACCAACAC GATGGCAGGT GAAAACATGG 8040

CTGCTTCTGC TAACAAGATG TCTGATACGA TGATGTCAGA GGATAAAGCT ATGCTACCAA 8100

ATACTGGTGA GACTCAAACA TCAATGGCAA GTATTGGTTT CCTTGGGCTT GCGCTTGCAG 8160

GTTTACTCGG TGGTCTAGGT TTGAAAAACA AAAAAGAAGA AAACTAATCA GCTAAGGAAA 8220

TAAATGATGG ATAGTGGGCT GACTAAGATT AGTTTAACAA CTCAATCAGC AATCAGGACT 8280

TTCTTTCAAT AGCAGATTAA AATCATCGTA AAACAATAAA AATAGTGTTA TACTTAAAGC 8340

AGTATAGCAC TGTTTTTATC AAAGGAGAGA CAGATGGGAA AGACAATTTT ACTCGTTGAC 8400

GACGAGGTAG AAATCACAGA TATTCATCAG AGATACTTAA TTCAGGCAGG TTATCAGGTC 8460

TTGGTAGCCC ATGATGGACT GGAAGCGCTA GAGCTGTTCA AGAAAAAACC GATTGATTTG 8520

ATTATCACAG ATGTCATGAT GCCTCGGATG GATGGTTATG ATTTAATCAG TGAGGTTCAA 8580

TACTTATCAC CAGAGCAGCC TTTCCTATTT ATTACTGCTA AGACCAGTGA ACAGGACAAG 8640

ATTTACGGCC TGAGCTTGGG AGCAGATGAT TTTATTGCTA AGCCTTTTAG CCCACGTGAG 8700

CTGGTTTTGC GTGTCCACAA TATTTTGCGC CGCCTTCATC GTGGGGGCGA AACAGAGCTG 8760

ATTTCCCTTG GCAATCTAAA AATGAATCAT AGTAGTCATG AAGTTCAAAT AGGAGAAGAA 8820

ATGCTGGATT TAACTGTTAA ATCATTTGAA TTGCTGTGGA TTTTAGCTAG TAATCCAGAG 8880

CGAGTTTTCT CCAAGACAGA CCTCTATGAA AAGATCTGGA AAGAAGACTA CGTGGATGAC 8940

ACCAATACCT TGAATGTGCA TATCCATGCT CTTCGACAGG AGCTGGCAAA ATATAGTAGT 9000

GACCAAACTC CCACTATTAA GACAGTTTGG GGGTTGGGAT ATAAGATAGA GAAACCGAGA 9060

GGACAAACAT GAAACTAAAA AGTTATATTT TGGTTGGATA TATTATTTCA ACCCTCTTAA 9120

CCATTTTGGT TGTTTTTTGG GCTGTTCAAA AAATGCTGAT TGCGAAAGGC GAGATTTACT 9180

TTTTGCTTGG GATGACCATC GTTGCCAGCC TTGTCGGTGC TGGGATTAGT CTCTTTCTCC 9240

TATTGCCAGT CTTTACGTCG TTGGGCAAAC TCAAGGAGCA TGCCAAGCGG GTAGCGGCCA 9300

AGGATTTTCC TTCAAATTTG GAGGTTCAAG GTCCTGTAGA ATTTGAGCAA TTAGGGCAAA 9360

CTTTTAATGA GATGTCCCAT GATTTGCAGG TAAGCTTTGA TTCCTTGGAA GAAAGCGAAC 9420

GAGAAAAGGG CTTGATGATT GCCCAGTTGT CGCATGATAT TAAGACTCCT ATCACTTCGA 9480

TCCAAGCGAC GGTAGAAGGG ATTTTGGATG GGATTATCAA GGAGTCGGAG CAAGCTCATT 9540

ATCTAGCAAC CATTGGACGC CAGACGGAGA GGCTCAATAA ACTGGTTGAG GAGTTGAATT 9600

TTTTGACCCT AAACACAGCT AGAAATCAGG TGGAAACTAC CAGTAAAGAC AGTATTTTTC 9660

TGGACAAGCT CTTAATTGAG TGCATGAGTG AATTTCAGTT TTTGATTGAG CAGGAGAGAA 9720

GAGATGTCCA CTTGCAGGTA ATCCCAGAGT CTGCCCGGAT TGAGGGAGAT TATGCTAAGC 9780

TTTCTCGTAT CTTGGTGAAT CTGGTCGATA ACGCTTTTAA ATATTCTGCT CCAGGAACCA 9840

AGCTGGAAGT GGTGGCTAAG CTGGAGAAGG ACCAGCTTTC AATCAGTGTG ACCGATGAAG 9900

GGCAGGGTAT TGCCCCAGAG GATTTGGAAA ATATTTTCAA ACGCCTTTAT CGTGTCGAAA 9960

CTTCGCGTAA CATGAAGACA GGTGGTCATG GATTAGGACT TGCGATTGCG CGTGAATTGG 10020

CCCATCAATT GGGTGGGGAA ATCACAGTCA GCAGCCAGTA CGGTCTAGGA AGTACCTTTA 10080

CCCTCGTTCT CAACCTCTCT GGTAGTGAAA ATAAAGCCTA AAACCCCTTT ACAAATCCAG 10140

CTATTCATGG TAGAATAGAT TTTGTGTGAA ATATCAGCAG GAAAGCATGA AGCTCGTCAA 10200

CAGGTGTCTT ATGACAAGTA ACCTTGGCTG TTTAGGCGAA GGGCATCTGC ACGG 10254

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9769 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CCGGCGACTA TCGATAACAC TTGACTTGGT AGCCCCACAT TTTGGACAAC GCATCCTTTC 60

CCTCCTTATC GTTTTCTTTT CATTATACCA TTTTTTAAGC GATTCCCAAA ACAATTCTTC 120

TTTTTGCTTG ACAAGTTTTT TGTTTTGTTG TATTATTTAA TTAAGACAAC AAGGTAAAAG 180

AAAGGAGACT AAGATGTCCT GGACATTTGA CAACAAAAAA CCCATCTATT TACAGATTAT 240

GGAGAAAATC AAGCTTCAGA TTGTTTCCCA TACACTGGAA CCCAATCAAC AACTTCCAAC 300

CGTGAGGAGC TAGCTAGCGA GGCTGGTGTC AATCCCAATA CCATCCAAAG AGCCTTATCA 360

GACCTTGAAC GAGAAGGATT TGTCTACAGC AAGCGAACAA CTGGACGATT TGTGACTAAG 420

GATAAGGAGC TAATCGCCCA GTCACGCAAA CAATTATCAG AAGAAGAATT GGAACACTTC 480

GTTTCCTCCA TGACCCATTT TGGCTATGAA AAAGAAGAAC TACCAGGCGT AGTCAGTGAT 540

TATATTAAAG GAGTTTAAGC CTATGTCATT ACTAGTATTT GAAAATGTAT CCAAATCATA 600

TGGAGCAACA CCAGCCCTTG AAAATGTTTC TCTTGACATT CCAGCTGGAA AAATTGTCGG 660

CCTTCTTGGG CCAAACGGCT CAGGAAAAAC AACCCTGATT AAACTAATTA ATGGCCTCTT 720

ACAACCAGAT CAAGGACGTG TCCTCATCAA CGACATGGAC CCAAGCCCAG CAACCAAGGC 780

CGTTGTAGCT TATTTGCCTG ATACGACCTA TCTCAATGAG CAAATGAAGG TCAAAGAAGC 840

CCTAACCTAC TTCAAGACCT TCTATAAAGA TTGTCAGATC TTGAACGCGC CCATCATCTA 900

CTTGCAGACC TGGGCATTGA TGAAAATAGT CGTCTCAAGA AACTATCAAA AGGAAACAAA 960

GAAAAGGTTC AACTGATTTT GGTTATGAGC CGTGATGCTC GTCTCTATGT TTTGGACGAA 1020

CCCATTGGTG GGGTGGATCC AGCAGCCCGT GCTTATATCC TCAATACCAT TATCAACAAC 1080

TACTCACCAA CTTCTACCGT TTTGATTTCT ACCCACTTGA TTTCTGATAT CGAGCCAATC 1140

TTGGATGAAA TTGTCTTCCT AAAAGACGGA AAAGTCGTCC GTCAAGGAAA TGTAGATGAT 1200

ATTCGCTACG AGTCAGGTGA ATCCATTGAC CAACTCTTCC GTCAGaATTT AAGGCCTAAG 1260

CAAAGGAGAT TATTTATGTT TTGGAATTTA GTTCGCTACG AATTTAAAAA TGTTAACAAG 1320

TGGTATTTAG CCCTCTACGC AGCCGTGCTA GTCCTTTCTG CCCTCATCGG AATACAGACA 1380

CAAGGCTTTA AAAATCTACC TTACCAAGAA AGTCAGGCTA CTATGCTACT TTTTCTAGCT 1440

ACAGTCTTTG GTGGCTTGAT GCTTACACTT GGGATTTCAA CCATTTTCTT GATTATTAAA 1500

CGCTTCAAAG GTAGTGTCTA CGACCGACAA GGCTATCTGA CTTTGACCTT GCCAGTTTCT 1560

GAACACCATA TCATCACAGC CAAACTAATC GGTGCCTTTA TCTGGTCATT GATTAGCACC 1620

GCTGTATTGG CTCTAAGTGC TGTTATTATT CTGGCTTTAA CAGCTCCAGA ATGGATTCCT 1680

CTTTCTTATG TGATTACATT TGTAGAAACA CATCTCCCTC AGATCTTTCT TACAGGTATA 1740

TCCTTCCTAC TAAATACTAT TTCAGGAATC CTCTGCATCT ACCTGGCTAT TTCCATTGGA 1800

CAGCTTTTCA ATGAATACCG TACAGCACTC GCTGTTGCAG TCTACATTGG TATCCAAATC 1860

GTCATTGGAT TTATTGAACT TTTCTTCAAT CTTAGTTCTA ATTTCTATGT CAATTCACTG 1920

GTAGGACTCA ATGACCATTT CTATATGGGA GCAGGTATAG CCATTGTTGA AGAACTCATA 1980

TTCATAGCTA TCTTTTATCT CGGAACCTAC TACATCTTGA GAAATAAGGT TAATTTGCTT 2040

TAAATAATTT TTACCTAGAT ATGTAACATA CTCATAGAAC AAAAGAGACC AGGCAAAAAG 2100

TCTTTAAAAT TAGAAAACGC ATAGTATCAG GTGTTGAATA TGTACTGCcC CCCAAAAGTT 2160

AGATTTTTTC TGTCTAACTT TTGGGGGCAG TTCATAAGAA CCTTGGTAAT ATGCGTTTTT 2220

TGTGAGCTGA CTTATTTCCT TTCACTATAT CGCAAAATGA AATAAGAACG GAACGATGGG 2280

ATTTTGGAAT TCAAATCAAT TTATAAGAAT GTTTTAGAAG TAATATTATC CTATTCCAGA 2340

TTCAGTTCAC TATACAATTG AGTTTTCAAG CAACCTGTTT ACATAATGTG TACATAATTA 2400

GGTTCGTGAT TCCACCCTTT TCACCTTTAA AAACCTCGCT TTCGCAAGGC TCTTCTATTT 2460

ATAAGATAAG GCACGTTTAA AGGTTTTCCA AATCCCTAAA TCATCCGTTT GAAGAACGAG 2520

ACTAGCATAC ATGCGTCCGA TAAATCCTGT TGCTACCACC GCAAAAATCA CTGTAATAGC 2580

AAGTGAAATC CATGCTTCTG CTCCCCCCGC ATAGTCATTA ATCGTTCGAA ACGGCATAAA 2640

GAAGGTCGAA ATAAAGGGAA TATAAGAACC AATCTTCAAG AGGAGATTGT CACCAGCTGC 2700

ACCTAGAGCT GTCACTCCAA AAAAACCACC CATAATCAAA ATCATCAAAG GCGACAAGGC 2760

TTTCCCTGAG TCCTCAGGAC GAGAAACCAT AGATCCTAGG AAGGCTGCCA AGACTACGTA 2820

CATGAAAAGA CTGATCAAAA TAAAGAGCAA GGTATTCAGT GAGATAGCAT CTCCCAAGTG 2880

ATCCAAAATA CCAGACTGAG CCAAGAATGG CAAATCTTTA AAGAGCAAAA CGGCAGCCAG 2940

ACCACCTACA ACATAGATCC CAATATGCGT TAAAATCACT AGAAACAGAG CCATCATCCG 3000

CGCATAGAAA TAGTGACTTG CCCTTATGCT AGAAAAAACG ACTTCCATAA TTTTGGTGCC 3060

TTTTTCACTG GCAACTTCCT GAGCTGTTAC ACCCGCATAG GTAATCAGAA TCATATAAAG 3120

AAAGAATCCT AAGGCACCTG CTGCAATTGT TTGAATAAAC TTTTTATTTT CCTTGGCTTC 3180

ATCAATCTTT TCTGTGAATT GAATTGTCTG CGCTAAGCGT TTTTCCTGCT CTTGAGACAA 3240

GGAAGCAGTT GAACGATTAA GCTGATTTTG CAGTTCATTG AGTGTACCTG TAACCTCAAA 3300

TTTAATTCCA TTTTCAAGCG ATGTTTCGCC ATGATAAACT GCCTTTAGAA CACTATCTTC 3360

TTGATCAATG GTCAAATAAC CTTTTAATTT TTCTTCTTTA ATTGCTTCTT TGGCACTTGC 3420

TTCGTCTTTA TAGTCGAAGT TAACACCATT TACATTCTTC AGTCCTTCTG CTACAGATGG 3480

CACTGTTGTC ACTACTGCCA CTTTATTATT TTTAGCCATA GAAGAACCTT GGAGATGCCC 3540

AATTCCTACA GAGATTCCTA AAAAGAGGAA CGGCGAAATC ACCATAAAGA AGAAACTCCA 3600

TGACTCGACA TGTCGAAGAT AGGTTTCCTT GATTACAACC CACATATTTC TCATACTTCC 3660

ACTCCTGATT CTAGTTTAAA GATTTCATCG ATAGTTGGCG CTTGTTGGTC AAATGTTGCG 3720

ATATATTGAC CTTGAGTCAA GATTGAGAAG AGTTCCCTTC CAGCGCTCTC ATCCTCCAAA 3780

ATCAATTTCC AACTGCCTTG TTTGGTCAAG CTCACCTGTT TGACATGAGG AAGATTTTCC 3840

AATTCTTCCT TGCTTCGTTC ACTTGAAACA AAGAGACGCG TTTTCCCGTA TTGATTGCGG 3900

ACATCCTGAA CTGGTCCGTG CAAGACCACA CGGCCATCTC GGATCATCAG AATATCGTCA 3960

CAAAGTTCCT CAACATTGGT CATGACATGG TCAGAAAAGA TAATGGTTGT CCGCGCTCTT 4020

TTTCCTGAAA AATGACTTGT TTGAGCAATT CTGTATTAAC TGGGTCCAAT CCACTAAAAG 4080

GCTCATCCAA GATAATCAGG TCTGGTTCAT GAATCAGAGT AATAATGAGC TGAATCTTCT 4140

GCTGATTTCC TTTTGACAGA CTCTTGATTT TATCTGTCAG CTTTCCTTTC ACTTCCAACC 4200

TCTTCATCCA TTGAGGGAGT TTTTCTTTGA CTTCTTTGGC ATCCATGCCT TTTAGAGTCG 4260

CCAAGTAGCG AACTTGTTCA AGAACTGTCA ATTTAGGCAT GAGATGCGTT CTTCAGGCAG 4320

ATAACCAATC CGAGCATAGG TCTCCTGACG AATATCCTGA CCATCCAGAC CGATTTCTCC 4380

CTGATATTCT AGGAATTTCA AAATACTATG GAAAATCGTT GTTTTTCCAG CACCATTTTT 4440

TCCGACTAGT CCCAAAATAC GACCTGGTCG CGCTTGAAAG TCAATACCAA ACAAAACTTG 4500

CTTGGATCCA AAACTTTTCT CTAGACTTCT TACTTCTAGC ATCTTTCACC TCCGAAATTT 4560

CTTGCACTCA TTATACTCCT TTTTGATAGC CTTTACAATG TTTTTTGTCC ATTTTTAGAA 4620

GACTATTGCT GTGTAAAATA TGGCCTGGAG CACTTTTATA CTCAATGAAA ATCAAAGAGC 4680

AAACTAGGAA GCTAGCCGTA GACTGCTCAA AGTACAGCTT TGAGGTTGCA GATAAAACTG 4740

ACGAAGTCgA CTCAAAACAC TGTTTTGAGG TTGTGGATAG AACTGACGAA kCrTAaCTAT 4800

ATCTACGGCA AGGCGAAcTG ACGTGGTTTG AAGAGATTTT CGAAGAGTAT TAGTGATAAA 4860

TCCATTATAC AGCAGCAAAC TTAATTTATA CCTTCCGCTC CTCAACTGTC TATTTTTAAT 4920

CCTGAATTGT TATTTGAGTA ACTCCTTTTT CCTCGTAAAG TTTTCTTCCT CTAAAACTTC 4980

TGGAAAAAGG CTAATAGTTT CAGACAACAT TTTTATAAGA AACAAGTTCA TCTGTCATTT 5040

CAAGAAGGAG TAATCCTTTA TCTACTAATG GACGGAACAG AATTCAACCG CTTGTCCGAT 5100

ATGTTTTCTA AGGATTATAT AGTAAAATGA AATAAGAACA GGACAAATTG ATCAGGACAG 5160

TCAAATTGAT TTCTAACAAT GTTTTAGAAG TAGATGTATA CTATTCTAGT TTCAATCTGC 5220

TATATCTATT ATGCACACCC CTATAGGATC TAATGAAAAT CACAACAGGC TCATTCATAG 5280

ATGGTTACCT AAGCCTAAGG GAACTAAGAA AACGACTACC AAGGAAGTCG CATTCATCGA 5340

AAAGTAGATT AACAACTATC CTAAAAAATG CTTGAACTAC AAGTCCCCCA GAGAAGACTT 5400

CTGGATGACT AACTTGAACT TGAAATTTAG CAATAATTAA TTCACTATCT AACTATATTT 5460

AGTAATTATT TCAGAACTGA TTAATATTAA AATTAACTAA CAATTCAAAG GATTCATACT 5520

AGCCATAAAT TACGTCCATC AGAGAGAGAC TCTTACTACT TTTAGATTTT AGTCTTTCTA 5580

GCTTCAGAAT ACATCTAAAC TTTAGGGAAA ATGACTATTC GAAAGCGCGA ATGCCTCAAA 5640

ATTATCTCAG ATAAGCTATT CGAAACTTAG AATGCTTTTA AATTTATGGA ATTGCGATTA 5700

TTCGAAACCT AGAATGCATA TAACCTTTAG TTGACAGACC TATTCTAAGT CTCGAAGGGC 5760

TATTTACTTT CTATTCCTTA TCAAAAAAGA CTCATTCCCC CTTTCTCCTC CAAAATATGG 5820

TATAGTAGAA ATATACTATC TATGAGGAGT TTACATGTCA CAGGATAAAC AAATGAAAGC 5880

TGTTTCTCCC CTTCTGCAGC GAGTTATCAA TATCTCATCG ATTGTCGGTG GGGTTGGGAG 5940

TTTGATTTTC TGTATTTGGG CTTATCAGGC TGGGATTTTA CAATCCAAGG AAACCCTCTC 6000

TGCCTTTATC CAGCAGGCAG GCATCTGGGG TCCACCTCTC TTTATCTTTT TACAGATTTT 6060

ACAGACTGTC GTCCCTATCA TTCCAGGGGC CTTGACCTCG GTGGCTGGGG TCTTTATCTA 6120

CGGGCACATC ATCGGGACTA TCTACAACTA TATCGGCATC GTGATTGGCT GTGCCATTAT 6180

CTTTTATCTA GTGCGCCTAT ACGGAGCTGC CTTTGTCCAG TCTGTCGTCA GCAAGCGCAC 6240

CTACGACAAG TACATCGACT GGCTAGATAA GGGCAATCGT TTTGACCGCT TCTTTATTTT 6300

TATGATGATT TGGCCCATTA GCCCAGCTGA CTTTCTCTGT ATGCTGGCTG CCCTGACCAA 6360

GATGAGCTTC AAGCGCTACA TGACCATCAT CATTCTGACC AAACCCTTTA CCCTCGTGGT 6420

TTATACCTAC GGTCTGACCT ATATTATTGA CTTTTTCTGG CAAATGCTTT GACACGTAAA 6480

AAATCCGTTT GGTTTCCCAA GTGGATTTTT AAAGCGTAGA TTAACTATAG CTTGATACTA 6540

AATATACTTT GGTATGGAAA TCATGCATAT TTTTCGATAG TGAGGCGAGG ACTTACCTAG 6600

CCTTTCCGCC GTGATAGAAA CACCTGAAAT CTAATGGTTT CAGGTATTCG GAAACTTTGA 6660

GCCTAGTGTC TCAAAGTTTA GGTATGGAAT TTTGAAGAAA GTCGCTACCG TCCGTAATCA 6720

CTTAAGGAAA GGCTCAAAAA TATTGTTTTC AACCACAAAA TCCGTTTGGT TTCCCAAGCG 6780

GATTTTGTGC TTTATTTTGA AACTTCTTTT GCAAGAACAA AGTTCCCAAG TGTGGCAGAA 6840

CCATTTCCTG CGACTGCTGG CGTCACGATA TAGTCACGCA CATCTGGTAC TGGTAGGTAA 6900

CCATTAAGAA GAGATGTAAA TTTCTCACGG ACACGGTCCA GCATATGTTG TTGAGCCATG 6960

ACCCCTCCAC CAAAGACAAT CACGTCTGGG CGGAAAGTCA CTGTCGCATT AACCGCAGCT 7020

TGAGCGATAT AGTAGGCTTG AACATCCCAA ACAGGGTTGT TGAGTTCAAT AGTTTCCCCA 7080

CGTACACCTG TACGAGCTTC CAAACTTGGA CCAGCTGCAT AACCTTCTAG ACATCCCTTA 7140

TGGAAAGGAC AAACACCCTT AAACTCTTTT TCAATATCCA TTGGGTGTCT AGCAACATAA 7200

TAATGACCCA TTTCAGGGTG ACCCACACCA CCGATAAACT CACCACGTTG GATGACGCCT 7260

GCACCGATAC CTGTACCGAT TGTGTAGTAA ACCAAGTTTT CGATACGACC ACCAGCATTG 7320

TTACGGGCAA CCATTTCACC GTAAGCAGAG CTGTTTACGT CTGTTGTGAA GTACATTGGC 7380

ACGTTTAGGG CGCGACGAAG GGCACCAAGC AAGTCTACAT TTGCCCAGTT TGGTTTTGGA 7440

GTCGTCGTGA TAAAGCCATA AGTTTTTGAG TTTTTGTCAA TATCAATCGG CCCAAATGAA 7500

CCAACTGCAA GACCAGCAAG GTTATCGAAT TTTGAGAAGA ACTCAATGGT TTTATCGATT 7560

GTTTCGATTG GAGTTGTTGT TGGAAATTGT GTTTTTTCTA CAACGTTAAA GTTTTCATCA 7620

CCGACAGCAC AGACAAACTT TGTACCGCCC GCTTCCAAGC TTCCATATAA TTTTGTCATG 7680

ATAAACCTCT TGTTTTTATT TTCTTTATTA TAGCATACTT CGAAAGTCTA AATGTCTCTA 7740

TTTTTTAGAT TTTCCTCTGT AAATCTTACT ATCTAATAAA AACGAACAAA CATGTCATTT 7800

GTTCGTTTTC ACATTAGAGA GGATTGATTA GATTTTCACT TCGATCACAG CATCCCCCTT 7860

AGCAACTGAA CCTGTTGCGA CTGGAGCTAC TGAAGCGTAG TCACCTGTAT TTGTAACGAT 7920

AACCATTGTT GTATCATCAA GTCCAGCTGC AGCGATTTTG TTTGAGTCAA ATGTTCCAAG 7980

AACATCGCCA GCTTTCACCT TATTACCTTG AGCAACTTTT GTTTCAAAAC CGTCACCGTT 8040

CATAGATACA GTATCAATAC CAACATGAAT CAAAACTTCA GCACCATTTC TTGTTTTCAA 8100

ACCAAAAGCG TGCCCTGTTG GAAAGGCAAT TGAAACTTCA GCATCAGCTG GTGCATAGAC 8160

CACGCCTTGG CTTGGTTTCA CAACGATACC TTGTCCCATA GCTCCACTTG AGAAGACTGG 8220

GTCATTGACA TCAGCAAGAG CGACAACATC ACCGACGATA GGAGTTACAA GTGTTTCATT 8280

TTGAAGAGCT GCTGGCGCAA CTTCTTCTTT TTCTTCAGCC ACTTCAGCTC GTTTTGCAGC 8340

TGCAGTTGCG TCTACTTCAT CTTCGTAACC AAACATGTAA GTAAGAGCAA AACCAAGGGC 8400

AAATGATACA GCTACCATAA GAAGGTATTG TGGAAGTTGT CCGTTACCAA CATAAAGCAT 8460

TGTACCAGGG ATGATGGTGA TACCATTACC AGTACCAGCA AGTCCAAGGA TAGAAGCCAA 8520

TCCACCACCG ATTGCACCAG CAATCAATGA AAGGAAGAAT GGTTTACGGA AGCGCAAGTT 8580

CACCCCGAAG ATAGCAGGCT GTGTAATACC TAGGAAGGCA GAAAGAGCAG CCGGGAAAGC 8640

AAGTGTTTTC AGTTTTGGAT TTTTTGTTTT AACACCAACC GCAACAGTAG CAGCACCTTG 8700

AGCTGTCATA GCAGCTGTGA TGATAGCGTT GAATGGGTTA GCATGGTCAG CAGCAAGTAA 8760

TTGCACTTCA AGCAAGTTGA AGATGTGGTG CACACCTGAC ACGACGATCA ATTGGTGAAC 8820

CCCACCAATC AAGAAACCAC CAAGACCAAA TGGCATGCTA AGAATCGCTT TTGTAGCAAT 8880

AAGGATGTAG TTTTCAACAA CGTGGAAAAC TGGTCCAATG ACAAAGAGTC CAAGGATAGA 8940

CATGACCAAA AGTGTCACGA ATGGTGTTAC CAAGAGGTCA ATGACATCTG GAACAACTTG 9000

CGGACAGCTT TTTCAAATTT AGCTCCGACA ACCCCGATGA TGAAGGCTGG AAGAACGGAA 9060

CCTTGCAAAC CAACAACAGG GATGAAACCA AAGAAGTTCA TCGCTGTTAC TTCACCACCT 9120

TGAGCAACTG CCCAAGCGTT TGGAAGTGAG CCAGAGACAA GCATCATACC AAGAACGATA 9180

CCAACGGCAG GATTTCCACC AAATACACGG AAGGTTGACC ACACAACCAA ACCTGGCAAG 9240

ATGATGAAGG CTGTATCTGT CAAGATTTGT GTGTAAGTTG CAAAGTGACC TGGAAGTGGC 9300

ATTTCAAGAG CGTTGAAAAG ACCACGCACA CCCATGAAGA GACCTGTCGC TACGATAACT 9360

GGGATGATTG GAACGAAAAC ATCACCAAAA GTACGGATAG CACGTTGGAA CCAGTTCCCT 9420

TGTTTAGCAA CTTCTGCTTT CATGTCATCC TTAGATGATG TTGGTAATCC AAGTACAACA 9480

ACTTCATCGT ACATTTTGTT AACTGTACCT GTACCAAAGA TAATTTGGTA TTGCCCTGAG 9540

TTAAAGAAAG CACCTTGAAC TTTTTCCAAG TTCTCAATCA CTTCTTTATT GATTTTCTCT 9600

TCATCTTTGA CCATGACACG TAGACGAGTC GCACAGTGGG CAACACTATT GACATTTTCA 9660

CGTCCGCCCA AGGCATCGAT GACTTTTTTT GCAATTTCCT GATTGTTCAT TTGCAAAAAT 9720

CTCCTTATAT AACATTTTGT TCTTGTTTGA AAGCGATTTT ATTCGCCGG 9769

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3149 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CGCTTGAGTG CTAATTCATA GTTCTATTGT ATCACTTGGT CAGAAATAAT CAAGAAAAAA 60

GTCTGACTTT CTCAAGATAA AAAGCCTGAG ACCAACTCAG ACTTTTTAAT TCTTAAAATG 120

GCAATTCTTC CTCTTCCAAG ACCAAATCTG CCAAATCTTG GCCTGCATTA TTTTCACGCA 180

TAGCACGTTG GGCACGACTT TCCAAGAGTT GGAATCCTGT GACAAGTACT TCGGTCACGT 240

AGTTCATTTG GCCATTTTTC TCAAAGCGAC GGGTACGCAA TTCTCCATCA ACGGAAATGA 300

GACTACCTTT GGTTGCGTAC TTGCCAAAGT TTCTGCTAGT CTGCCCCATA GGACCATATT 360

GACAAAATCA GCTTCACGTT CACCGTTTTG GTCTTTGTAA CGACGGTTCA CAGCGATAGT 420

TGCTCGCGCT ACCGACTTGT CATTGTTGGT TTTGTGCAAT TCTGGTGTAG ACGTTAAACG 480

TCCAATCAAG ATAACTTTAT TATACATATT TTCTTCCTCC TACTTATCTA TTCGTAGGAA 540

ATCAAAAAAA GTTACAGAAA TTTGTAACTT TTCGAGAAAA TTTTTTATTT TTTATGAACC 600

ATGAAACCTG TCGCCTGTTG ATTGGCCATA ATGGTCATAT CTGTAATCTG AACACGACGA 660

GGTTGACTAG TCACATAGAC TACTGTATCT GCAATATCCT GAGCTTGCAA AGCTTCTATT 720

CCTTGGTAAA CGGACGCAGC TCGTTCTTTA TCACCATGAA AACGCACTGT AGAAAAATCT 780

GTTTCGACAA TTCCAGGCTG AATGGTCGTC ACCTTGATAT CCGTTGCGAT GGTATCAATT 840

CGCAGTCCAT CTGAAAAGGT CTTAACTGCC GCCTTGGTGG CTGAGTAAAC AGCTGCACCA 900

GCATAGGCAT AAATTCCTGC GGTTGACCCC ATATTGATAA TATGACCTTG ATTGGCTTTT 960

ACCATTGCTG GCAAGAAACA GCGAGTGACT GCCATCAAAC CTTTGACATT GGTATCCAAC 1020

ATGGTCAGCA TATCCAACTC TTCATAGTCT TGATAGGGAG CTAAGCCAAG AGCCAGTCCT 1080

GCGTTATTGA CCAGGATGTC AATCTGACCT ATCGTTTCTA AAATATCAGA GCAGACAGTC 1140

TTTACCATTG TCATATCCGT GACATCTAGG AGAAAAGTCC AAACTGTTTG ATTTGGAAAA 1200

GTTTCTGCAA ACTCCGCCTT AAGAGCTTCT AGTCTGTCTA TCCGTCGTCC TGTTAGAACG 1260

ACATCCTCAC CCTGCTCCAG ATAAGCACGC GCAATCGCTT CACCGATTCC TGATGTCGCT 1320

CCTGTAATCA CAACATTTTT TGCCATCTTA TTTCCTTCTA GCTGGTCTAT CAGATATTAA 1380

CAACTTCTTA GGCAGTCCAG TGTTTCGCTG GGTCGAACGG TGTTCCGACA ACTTGGTCTT 1440

CTGATAATTC AAGCACCCCA CGTTTTTGTG GAGCATTTGG CAGATGCAAT TCACGAGGAC 1500

TGCACATCAT ACCAAAACTC TTTTCACCAC GAAGTTCACC TGGGAAAATG AGATTCCCTT 1560

TTGGCATCAT AGCTCCAGGA AGCGCGACAA TGGTTTTCAA CCCCACACGC GCATTGGGAG 1620

CTCCTGCAAC GATTTGTACA GTCTTATCAC TTGCGACTGC AACTTGGCAG ATGTTGAGGT 1680

GGTCACTATC TGGATGGGCT ACCATCTCAA CAATTTCACC TACAACAAAC TTAGGTTCCT 1740

TATCATTAAC AATTTCTTCT GTAAAACCTT CCGCCTGCAA CTCTTGGTTC AAACGAGCGA 1800

CTTGCTCATC TGTCAAAAAG ACTTGACCGC GCTCTGCAAT TTCAAATAAA CTTGAAACTT 1860

CGAAAATATT CCAAGCCACT GTTTCCCCAT TATCTTTGAG AAAAACACGG GCTACCTTGC 1920

CTTTGCGCTC CACATCCAGT TTGGCATCTC CGCTATTTTT CACGATGACC ATAAGGACAT 1980

CACCGACATG TTCTTTATTA TATGTAAAAA TCATTGTTTC CTTTTTCTCC TATTTCAGTC 2040

CTGCTAAAAA GTCATTGATT TGTTGCTTGC TTTTACGGTC GCGATTGACA AAACGACCGA 2100

TTTCCTTGTC CTTTTCTAGA ACAACAAGGC TAGGAATTCC GTAAACATCC CAGAGTTTGG 2160

CCAAATCCAT ATACTGATCT CGGTCCATTC GAATAAAGGT GAACTCTGGA TTGGTCTCCT 2220

CAATCTCTGG TAAGGCAGGA TAAATATAAC GACAATCGCT ACACCAGTCT GCCACAAAAA 2280

TGAAGACCTT CTTGCCCGCT TTTTCCACTA AAGATGCTAA TTCTTCTAAA CTTGCTGGCT 2340

GTATCATAAG ACTTCCTCCT CATAGACTAG GTCTTCATTT TCATAGACAA AGGTATAATG 2400

ACGGCCATCC TCAAAAATGA CGCCACCAAC CAAGCTCTCC AGACTGCTTT CGTAAACTTG 2460

AACATAAAGG GTCGCAATTT CCCCCATGTC GGAAAAATGG TCTCGCACAA TCTCTGTCAA 2520

CTCTTCCTGA GTCTTCATGA GCTTACGGTC ATCTGCAACT TTTTTCGTAG CAAGAGCAAG 2580

GCTTCCGATA CCTAGCAGAG CCAAGCCTGC CATCCACATT TTTTTAGCTT TCATACCATT 2640

CATTTTAACA CAAAAAAGGC TTCAGGACAA ATGAGGAAGC AGCAGAAAAG CAAGTAAAAA 2700

GCCTCTTCCT TTAAGGAAAA GGACTTCTTA TACTCAATGA AAATCAAAGA CCAAACTAGG 2760

AAGCTAGCCG CAGGCTGCTC AAAGCACTGC TTTGAGGTTG TAGATAGAAC TGACGAgTCa 2820

CTCAAAACAC TGTTTTGAGG TTGTGGATGA AGCTGACGTG GTTTGAAGAG ATTTTCGAAG 2880

AGTATTATTC TTATTGCCAG GCACCTAAGT TGCCAACGTA GTAACTATCA GGTGTGTAGG 2940

TATTGCGAGC ATCTTACCTG ATGAAGCCAG ATAATACTAC TTGCCATTGT CTTTGACCCA 3000

ATCATTCGCA ATCATGGAAC CAGAAGAACT TACATAATAC CATTCTCCCT TGTCATAAAC 3060

CCAAGTACTG ACTTTCATGG TTCCTGAGCA ATTAAAGGCA AAAAAACTGT CCAATAACAT 3120

TCGTTTTTTA AAAGCATTTG ACACTACAT 3149

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10240 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

CCAAAAAATTC AACCTTTAAG GGGAGTCCAG AGAGACTCAC AAGGTGTCAG ATAAAAGAAT 60

GGTGCAATTT TCTAGAGGAG ACTTTTTGAG TGTGCTCTCT TGTGTTGTAC GATTTTAACT 120

GAGGCCTTGC ACTAGCAAGG TCTTTTCTTT ATCTGGTCCC CTTAAAATTT AAGGAGGAAA 180

AGTTATGAAT CCCACATGTA AGAAGCGTTT GGGTGTCATT CGGTTGGAAA CCATGAAGGT 240

GGTTGCACAA GAGGAAATCG CGCCACAATC TTTGAATTAG TCCTAGAAGG AGAAATGGTT 300

GAAGCCATGC GAGCAGGCCA ATTTCTTCAT CTGCGTGTAC CGGACGATGC CCATCTCTTA 360

CGTCGTCCTA TTTCAATTTC GTCTATTGAC AAGGCAAACA AGCAGTGTCA CCTCATTTAT 420

CGGATTGACG GAGCTGGGAC TGCAATTTTT TCAACCTTAA GTCAGGGAGA CACTCTTGAT 480

GTGATGGGGC CTCAGGGAAA TGGTTTTGAC TTGTCTGACC TTGATGAGCA GAATCAGGTT 540

CTCCTTGTTG GTGGTGGGAT TGGTGTTCCA CCCTTGCTTG AGGTGGCCAA GGAATTGCAT 600

GAACGTGGAG TGAAAGTAGT GACAGTCCTC GGTTTTGCTA ATAAGGATGC TGTTATTTTG 660

AAAACGGAAT TGGCTCAGTA TGGTCAGGTC TTTGTAACGA CAGATGATGG TTCTTATGGC 720

ATCAAGGGAA ATGTTTCCGT TGTTATCAAT GATTTAGACA GTCAGTTTGA TGCTGTTTAC 780

TCGTGTGGGG CTCCAGGAAT GATGAAGTAT ATCAATCAAA CCTTTGATGA TCACCCAAGA 840

GCCTATTTAT CTCTGGAATC TCGTATGGCT TGTGGGATGG GAGCTTGCTA TGCCTGTGTT 900

CTAAAAGTAC CAGAAAACGA GACGGTCAGC CAACGCGTCT GTGAAGATGG TCCTGTTTTC 960

CGCACAGGAA CAGTTGTATT ATAAGGAGAA AATTATGACT ACAAATCGAT TACAAGTTTC 1020

TCTACCTGGT TTGGATTTGA AAAATCCGAT TATTCCAGCA TCAGGCTGTT TTGGCTTTGG 1080

ACAAGAGTAT GCCAAGTACT ATGATTTAGA CCTTTTAGGT TCTATTATGA TCAAGGCGAC 1140

AACCCTTGAA CCACGTTTTG GGAATCCAAC TCCAAGAGTG GCAGAGACGC CTGCTGGTAT 1200

GCTCAATGCA ATTGGCTTGC AAAATCCTGG TTTAGAGGTT GTTTTGGCTG AAAAGCTACC 1260

TTGGCTGGAA AGAGAATATC CAAATCTTCC TATTATTGCC AATGTAGCTG GTTTTTCAAA 1320

ACAAGAGTAT GCAGCTGTTT CTCATGGGAT TTCCAAGGCA ACTAATGTAA AAGCTATCGA 1380

GCTCAATATT TCTTGTCCCA ATGTTGACCA CTGTAATCAT GGACTTTTGA TTGGTCAAGA 1440

TCCAGATTTG GCTTATGATG TGGTGAAAGC AGCTGTGGAA GCCTCAGAAG TGCCAGTTTA 1500

TGTCAAATTA ACCCCGAGTG TGACCGATAT CGTTACTGTC GCAAAAGCTG CAGAAGATGC 1560

GGGAGCAAGT GGCTTGACCA TGATCAATAC TCTGGTTGGA ATGCGCTTTG ACCTCAAAAC 1620

TAGAAAACCA ATCTTGGCCA ATGGAACAGG TGGAATGTCT GGTCCAGCAG TCTTTCCAGT 1680

AGCCCTCAAA CTCATCCGCC AAGTTGCCCA AACAACAGAC CTGCCTATCA TTGGAATGGG 1740

AGGAGTGGAT TCGGCTGAAG CTGCCCTAGA AATGTATCTG GCTGGGGCAT CTGCTATCGG 1900

AGTTGGAACA GCTAACTTTA CCAATCCTTA TGCCTGCCCT GACATCATCG AAAATTTACC 1860

AAAAGTCATG GATAAATACG GTATTAGCAG TCTGGAAGAA CTCCGTCAGG AAGTAAAAGA 1920

GTCTCTGAGG TAAACTGCAA TCAATCTGTT CTTGATTTTT TATTAGTTTG TAATATGAAT 1980

TTAGGAGAAT TTTGGTACAA TAAAATAAAT AAGAACAGAG GAAGAAGGTT AATGAAGAAA 2040

GTAAGATTTA TTTTTTTAGC TCTGCTATTT TTCTTAGCTA GTCCAGAGGG TGCAATGGCT 2100

AGTGATGGTA CTTGGCAAGG AAAACAGTAT CTGAAAGAAG ATGGCAGTCA AGCAGCAAAT 2160

GAGTGGGTTT TTGATACTCA TTATCAATCT TGGTTCTATA TAAAAGCAGA TGCTAACTAT 2220

GCTGAAAATG AATGGCTAAA GCAAGGTGAC GACTATTTTT ACCTCAAATC TGGTGGCTAT 2280

ATGGCCAAAT CAGAATGGGT AGAAGACAAG GGAGCCTTTT ATTATCTTGA CCAAGATGGA 2340

AAGATGAAAA GAAATGCTTG GGTAGGAACT TCCTATGTTG GTGCAACAGG TGCCAAAGTA 2400

ATAGAAGACT GGGTCTATGA TTCTCAATAC GATGCTTGGT TTTATATCAA AGCAGATGGA 2460

CAGCACGCAG AGAAAGAATG GCTCCAAATT AAAGGGAAGG ACTATTATTT CAAATCCGGT 2520

GGTTATCTAC TGACAAGTCA GTGGATTAAT CAAGCTTATG TGAATGCTAG TGGTGCCAAA 2580

GTACAGCAAG GTTGGCTTTT TGACAAACAA TACCAATCTT GGTTTTACAT CAAAGAAAAT 2640

GGAAACTATG CTGATAAAGA ATGGATTTTC GAGAATGGTC ACTATTATTA TCTAAAATCC 2700

GGTGGyTACA TGGCAGCCAA TGAATGGATT TGGGATAAGG AATCTTGGTT TTATCTCAAA 2760

TyTGATGGGA AAATrGCTGA AAAAGAATGG GTCTACGATT CTCATAGTCA AGCTTGGTAC 2820

TACTTCAAAT CCGGTGGTTA CATGACAGCC AATGAATGGA TTTGGGATAA GGAATCTTGG 2880

TTTTACCTCA AATCTGATGG GAAAATAGCT GAAAAAGAAT GGGTCTACGA TTCTCATAGT 2940

CAAGCTTGGT ACTACTTCAA ATCTGGTGGC TACATGGCGA AAAATGAGAC AGTAGATGGT 3000

TATCAGCTTG GAAGCGATGG TAAATGGCTT GGAGGAAAAA CTACAAATGA AAATGCTGCT 3060

TACTATCAAG TAGTGCCTGT TACAGCCAAT GTTTATGATT CAGATGGTGA AAAGCTTTCC 3120

TATATATCGC AAGGTAGTGT CGTATGGCTA GATAAGGATA GAAAAAGTGA TGACAAGCGC 3180

TTGGCTATTA CTATTTCTGG TTTGTCAGGC TATATGAAAA CAGAAGATTT ACAAGCGCTA 3240

GATGCTAGTA AGGACTTTAT CCCTTATTAT GAGAGTGATG GCCACCGTTT TTATCACTAT 3300

GTGGCTCAGA ATGCTAGTAT CCCAGTAGCT TCTCATCTTT CTGATATGGA AGTAGGCAAG 3360

AAATATTATT CGGCAGATGG CCTGCATTTT GATGGTTTTA AGCTTGAGAA TCCCTTCCTT 3420

TTCAAAGATT TAACAGAGGC TACAAACTAC AGTGCTGAAG AATTGGATAA GGTATTTAGT 3480

TTGCTAAACA TTAACAATAG CCTTTTGGAG AACAAGGGCG CTACTTTTAA GGAAGCCGAA 3540

GAACATTACC ATATCAATGC TCTTTATCTC CTTGCCCATA GTGCCCTAGA AAGTAACTGG 3600

GGAAGAAGTA AAATTGCCAA AGATAAGAAT AATTTCTTTG GCATTACAGC CTATGATACG 3660

ACCCCTTACC TTTCTGCTAA GACATTTGAT GATGTGGATA AGGGAATTTT AGGTGCAACC 3720

AAGTGGATTA AGGAAAATTA TATCGATAGG GGAAGAACTT TCCTTGGAAA CAAGGCTTCT 3780

GGTATGAATG TGGAATATGC TTCAGACCCT TATTGGGGCG AAAAAATTGC TAGTGTGATG 3840

ATGAAAATCA ATGAGAAGCT AGGTGGCAAA GATTAGTACT ATAAGTGAAT ATGATTTGAG 3900

TGAATAGTAA GTTAAAAATC CTGATTTCAA GTAAAATCAG GATTTTTTCA TGGATGCAAT 3960

TTTTTTGGAG TCTGGTGTGA CGCGGAGGGT CTTTTGTCCT GTGTAAGTGA CAAAGCCGGG 4020

TTTTCCACCA GTTGGTTTAT TGAGTTTTTT GACTTCAATC ATATCTACCT GCACCAGATT 4080

CGACAGGCGC CCTTGAGAGA AGTAGGCAGC TAACTCTGCT GCGTCTGTCT TGACTGCATC 4140

AGATGGGTCA AGATTTCCTG AGATGACAAC ATGGCTTCCA GGAATGTCCT TAGCATGGAA 4200

CCAAAGTTCC TCCTTGCGGG CCATTTTAAA GGTCAATTCC TCATTTTGAA GATTGTTTCG 4260

TCCGACATAG ATGATGGTTT TGCCATCGCT TGCTAGATAT TGTTCTAGTT TTTTGCGTTT 4320

CTGGATTTTC TCCCGTTGTC TTCTGCGGAT AAAACCTGTT TGAATCAATT CTTCACGGAT 4380

TTCAGCGATT TCTTCCAGTC CAGCTTGGTT GAGGACGGTT TCTACACTTT CCAGATAGAG 4440

AATAGTGGCT TTGGTTTCTT CAATCAAATC AGTCAAGTAT TTGACAGCTT CTTTGAGTTT 4500

CTGATACCGT TTAAAATAGC GTTGGGCATT CTGGTTGGGA GTCAGAGCCT TATCAAGCGC 4560

AATCATGATA GGTTGGTTGG TATAGTAGTT GTCTAGGATA ACCTGGTCTT GGTCGTTAGG 4620

CACTTGGTGG AGGAAGGTTG TCAGCAATTC TCCTTTTTGA CGAAATTCTT CAGCGTTGTC 4680

TGTCGCCAGT AACTCTTTTT CCTGTTTTTT GAGTTTGTGT CGGTTTTTCT GAAGTTCATT 4740

TTCAACACGA CGAATCAGTT CACTGGCCTG CTGTTTGACG CGGTCGCGCT CAGCCTTATC 4800

CTTATAGTAG GTGTCCAACA AATCAGAAAG ATTTGCAAAA GGCTCTCCCA CCTGATTTGC 4860

AAAAGGAACT GGACTGAAGG AAGTCTCAGT CAAGCATGGC TTGGTTTCTT GATTGAAAAA 4920

ATTTCGGAAA GCGGAAAGTT TTTCACTAAC CAGTATCCTT TCCAATTCAT TTGCCGTATC 4980

GCGTCCCAGA CCTTGAAAGA GGCTTTGAAG ATTTTTTGCT GTTAGTTCTT GGGTTTGCAG 5040

GATTTCAAAG AGCTTTTCAT CCTTGATAGT AAAAGGATTG AGAGATTTTG TACTTGGCGG 5100

AGCGATATAG GTCGATCCTG GAAGTAAGGT GCGGTAGCTA TTTTGTGAAA AGCCGACGTG 5160

TTTGATAACT TCGAGGATTT TATGACTGCT TTTATCGACC AGTAGAATAT TACTGTGTTT 5220

CCCCATAATT TCGATAATCA AGGTAGCCTG GATATGGTCT CCAATCTCGT TTTTATTGGA 5280

AACTGTAATT TCCACAATAC GGTCATTTTC CACTTGCTCA ATCGACTCAA TCAGGGCCCC 5340

CTGCAAATAC TTTCTCAAAA CCATGATAAA GGTAGAAGGT TGAGCTGGAT TTTCAAAAGT 5400

CGTTTGGGTC AGCTGAATGC GTCCAAAAAC TGGATGGGCA GAAAGGAGCA GGCGATGGCT 5460

TTGGCGATTG CTGCGGATTT GCAAGACCAA CTCTTGTTCA AAAGGCTGAT TGATTTTCTG 5520

GATGCGACCA TTCACTAATT CGCTTCGCAA TTCCTCAACT ATGTGGTGTA AAAAAAATCC 5580

GTCAAATGAC ATCGTTCTCT CCTTGTGATT GTATTCCATA GTATTATATC AAAAAGGTAG 5640

AATAAAATCA TGGAAATGTG GTATAATAAA GCCAAGTAAA GAGAAACGAG AAGCACATGT 5700

ATATTGAAAT GGTAGATGAA ACTGGTCAAG TTTCAAAAGA AATGTTGCAA CAAACCCAAG 5760

AAATTTTGGA ATTTGCAGCC CAAAAATTAG GAAAAGAAGA CAAGGAGATG GCAGTCACTT 5820

TTGTGACCAA TGAGCGTAGT CATGAACTTA ATCTGGAGTA CCGTAACACC GACCGTCCGA 5880

CAGATGTCAT CAGCCTTGAG TATAAACCAG AATTGGAAAT TGCCTTTGAC GAAGAGGATT 5940

TGCTTGAAAA TTCAGAATTG GCAGAGATGA TGTCTGAGTT TGATGCCTAT ATTGGGGAAT 6000

TGTTCATCTC TATCGATAAG GCTCATGAGC AGGCCGAAGA ATATGGTCAC AGCTTTGAGC 6060

GTGAGATGGG CTTCTTGGCA GTACACGGCT TTTTACATAT TAACGGCTAT GATCACTACA 6120

CTCCGGAAGA AGAAGCGGAG ATGTTCGGTT TACAAGAAGA AATTTTGACA GCCTATGGAC 6180

TCACAAGACA ATAAACGAAA ATGGAAAAAT CGTGACTTGA TATCCAGTTT AGAATTTGCT 6240

TTGACAGGTA TTTTTACTGC TATCAAGGAA GAACGCAATA TGCGAACACA CGCAGTGACG 6300

GCTCTAGTGG TCATCCTTGC AGGTTTTGTT TTTCAGGTGT CACGAATCGA ATGGCTCTTT 6360

CTCCTATTGA GTATTTTCTT GGTAGTAGCC TTTGAGATTA TCAACTCTGC TATTGAAAAT 6420

GTGGTGGATT TGGCCAGTCA CTATCACTTT TCCATGCTGG CTAAAAATGC CAAGGATATG 6480

GCGGCCGGCG CGGTATTAGT GGTTTCTCTT TTCGCAGCCT TAACAGGCGC ATTGATTTTT 6540

CTCCCACGAA TCTGGGATTT ATTATTTTAA ACAGTAAGAG GAAATTATGA CTTTTAAATC 6600

AGGCTTTGTA GCCATTTTAG GACGTCCCAA TGTTGGGAAG TCAACCTTTT TAAATCACGT 6660

TATGGGGCAA AAGATTGCCA TCATGAGTGA CAAGGCGCAG ACAACGCGCA ATAAAATCAT 6720

GGGAATTTAC ACGACTGATA AGGAGCAAAT TGTCTTTATC GACACACCAG GGATTCACAA 6780

GCCTAAAACA GCTCTCGGAG ATTTCATGGT TGAGTCTGCC TACAGTACCC TTCGCGAAGT 6840

GGACACTGTT CTTTTCATGG TGCCTGCTGA TGAAGCGCGT GGTAAGGGGG ACGATATGAT 6900

TATCGAGCGT CTCAAGGCTG CCAAGGTTCC TGTGATTTTG GTGGTGAATA AAATCGATAA 6960

GGTCCATCCA GACCAGCTCT TGTCTCAGAT TGATGACTTC CGTAATCAAA TGGACTTTAA 7020

GGAAATTGTT CCAATCTCAG CCCTTCAGGG AAATAACGTG TCTCGTCTAG TGGATATTTT 7080

GAGTGAAAAT CTGGATGAAG GTTTCCAATA TTTCCCGTCT GATCAAATCA CAGACCATCC 7140

AGAACGTTTC TTGGTTTCAG AAATGGTTCG CGAGAAAGTC TTGCACCTAA CTCGTGAAGA 7200

GATTCCGCAT TCTGTAGCAG TAGTTGTTGA CTCTATGAAA CGAGACGAAG AGACAGACAA 7260

GGTTCACATC CGTGCAACCA TCATGGTCGA GCGCGATAGC CAAAAAGGGA TTATCATCGG 7320

TAAAGGTGGC GCTATGCTTA AGAAAATCGG TAGCATGGCC CGTCGTGATA TCGAACTCAT 7380

GCTAGGAGAC AAGGTCTTCC TAGAAACCTG GGTCAAGGTC AAGAAAAACT GGCGCGATAA 7440

AAAGCTAGAT TTGGCTGACT TTGGCTATAA TGAAAGAGAA TACTAAGTAG AGGTAGGCTC 7500

ATGCCTGCTT CTTGTTTTTA CAGAAGGAGG ACTTATGCCT GAATTACCTG AGGTTGAAAC 7560

CGTTTGTCGT GGCTTAGAAA AATTGATTAT AGGAAAGAAG ATTTCGAGTA TAGAAATTCG 7620

CTACCCCAAG ATGATTAAGA CGGATTTGGA AGAGTTTCAA AGGGAATTGC CTAGTCAGAT 7680

TATCGAGTCA ATGGGACGTC GTGGAAAATA TTTGCTTTTT TATCTGACAG ACAAGGTCTT 7740

GATTTCCCAT TTGCGGATGG AGGGCAAGTA TTTTTACTAT CCAGACCAAG GACCTGAACG 7800

CAAGCATGCC CATGTTTTCT TTCATTTTGA AGATGGTGGC ACGCTTGTTT ATGAGGATGT 7860

TCGCAAGTTT GGAACCATGG AACTCTTGGT GCCTGACCTT TTAGACGTCT ACTTTATTTC 7920

TAAAAAATTA GGTCCTGAAC CAAGCGAACA AGACTTTGAT TTACAGGTCT TTCAATCTGC 7980

CCTTGCCAAG TCCAAAAAGC CTATCAAATC CCATCTCCTA GACCAGACCT TGGTAGCTGG 8040

ACTTGGCAAT ATCTATGTGG ATGAGGTTCT CTGGCGAGCT CAGGTTCATC CAGCTAGACC 8100

TTCCCAGACT TTGACAGCAG AAGAAGCGAC TGCCATTCAT GACCAGACCA TTGCTGTTTT 8160

GGGCCAGGCT GTTGAAAAAG GTGGCTCCAC CATTCGGACT TATACCAATG CCTTTGGGGA 8220

AGATGGAAGC ATGCAGGACT TTCATCAGGT CTATGATAAG ACTGGTCAAG AATGTGTACG 8280

CTGTGGTACC ATCATTGAGA AAATTCAACT AGGCGGACGT GGAACCCACT TTTGTCCAAA 8340

CTGTCAAAGG AGGGACTGAT GGGAAAAATC ATCGGAATCA CTGGGGGAAT TGCCTCTGGT 8400

AAGTCAACTG TGACAAATTT TCTAAGACAG CAAGGCTTTC AAGTAGTGGA TGCCGACGCA 8460

GTCGTCCACC AACTACAGAA ACCTGGTGGT CGTCTGTTTG AGGCTCTAGT ACAGCACTTT 8520

GGGCAAGAAA TCATTCTTGA AAACGGAGAA CTCAATCGCC CTCTCCTAGC TAGTCTCATC 8580

TTTTCAAATC CTGATGAACG AGAATGGTCT AAGCAAATTC AAGGGGAGAT TATCCGTGAG 8640

GAACTGGCTA CTTTGAGAGA ACAGTTGGCT CAGACAGAAG AGATTTTCTT CATGGATATT 8700

CCCCTACTTT TTGAGCAGGA CTACAGCGAT TGGTTTGCTG AGACTTGGTT GGTCTATGTG 8760

GACCGAGATG CCCAAGTGGA ACGCTTAATG AAAAGGGACC AGTTGTCCAA AGATGAAGCT 8820

GAGTCTCGTC TGGCAGCCCA GTGGCCTTTA GAAAAAAAGA AAGATTTGGC CAGCCAGGTT 8880

CTTGATAATA ATGGCAATCA GAACCAGCTT CTTAATCAAG TGCATATCCT TCTTGAGGGA 8940

GGTAGGCAAG ATGACAGAGA TTAACTGGAA GGATAATCTG CGCATTGCCT GGTTTGGTAA 9000

TTTTCTGACA GGAGCCAGTA TTTCTTTGGT TGTACCTTTT ATGCCCATCT TCGTGGAAAA 9060

TCTAGGTGTA GGGAGTCAGC AAGTCGCTTT TTATGCAGGC TTAGCAATTT CTGTCTCTGC 9120

TATTTCCGCG GCGCTCTTTT CTCCTATTTG GGGTATTCTT GCTGACAAAT ACGGCCGAAA 9180

ACCCATGATG ATTCGGGCAG GTCTTGCTAT GACTATCACT ATGGGAGGCT TGGCCTTTGT 9240

CCCAAATATC TATTGGTTAA TCTTTCTTCG TTTACTAAAC GGTGTATTTG CAGGTTTTGT 9300

TCCTAATGCA ACGGCACTGA TAGCCAGTCA GGTTCCAAAG GAGAAATCAG GCTCTGCCTT 9360

AGGTACTTTG TCTACAGGCG TAGTTGCAGG TACTCTAACT GGTCCCTTTA TTGGTGGCTT 9420

TATCGCAGAA TTATTTGGCA TTCGTACAGT TTTCTTACTG GTTGGTAGTT TTCTATTTTT 9480

AGCTGCTATT TTGACTATTT GCTTTATCAA GGAAGATTTT CAACCAGTAG CCAAGGAAAA 9540

GGCTATTCCA ACAAAGGAAT TATTTACCTC GGTTAAATAT CCCTATCTTT TGCTCAATCT 9600

CTTTTTAACC AGTTTTGTCA TCCAATTTTC AGCTCAATCG ATTGGCCCTA TTTTGGCTCT 9660

TTATGTACGC GACTTAGGGC AGACAGAGAA TCTTCTTTTT GTCTCTGGTT TGATTGTGTC 9720

CAGTATGGGC TTTTCCAGCA TGATGAGTGC AGGAGTCATG GGCAAGCTAG GTGACAAGGT 9780

GGGCAATCAT CGTCTCTTGG TTGTCGCCCA GTTTTATTCA GTCATCATCT ATCTCCTCTG 9840

TGCCAATGCC TCTAGCCCCC TTCAACTAGG ACTCTATCGT TTCCTCTTTG GATTGGGAAC 9900

CGGTGCCTTG ATTCCCGGGG TTAATGCCCT ACTCAGCAAA ATGACTCCCA AAGCCGGCAT 9960

TTCGAGGGTC TTTGCCTTCA ATCAGGTATT CTTTTATCTG GGAGGTGTTG TTGGTCCCAT 10020

GGCAGGTTCT GCAGTAGCAG GTCAATTTGG CTACCATGCT GTCTTTTATG CGACAAGCCT 10080

TTGTGTTGCC TTTAGTTGTC TCTTTAACCT GATTCAATTT CGAACATTAT TAAAAGTAAA 10140

GGAAATCTAG TGCGAGTAAA AATCAATCTC AAATGCTCCT CTTGTGGCAG TATCAATTAC 10200

CTAACCAGTA AAAATTCAAA AACCCATCCA GACAgATTGA 10240

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13206 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CGCTTTATCG TGGACGTGGT CAAGCCGAGA ATTTCATCAA GGAGATGAAG GAGGGATTTT 60

TTGGCGATAA AACGGATAGT TCAACCTTAA TCAAAAACGA AGTTCGTATG ATGATGAGCT 120

GTATCGCCTA CAATCTCTAT CTTTTTCTCA AACATCTAGC TGGAGGTGAC TTCCAAACTT 180

TAACAATCAA ACGCTTCCGC CATCTTTTTC TTCACGTGGT GGGAAAATGT GTTCGAACAG 240

GACGCAAGCA GCTCCTCAAA TTGTCTAGTC TCTATGCCTA TTCCGAATTG TTTTCAGCAC 300

TTTATTCTAG GATTAGAAAA GTCAACCTGA ATCTTCCTGT TCCTTATGAA CCACCTAGAA 360

GAAAAGCGTC GTTAATGATG CATTAAAGAA CAGTCGAGAT GAAAAAATCG TGTGACGCAC 420

CAAGGGAGGA GTCTGCCCTT TTGAGGAAAT CTAGCGAGGA AAAACGATAC TGGAACAGCA 480

GAAAGTAAAA CTGACCTCAT GAGGAGGAAG AAAGTGGCTC ATGAGGTCAG GGGTTTTGTA 540

AGTTACATCT AGTTGAGAGA GGTATGAATG ATTTGGGATT AATCATTTCT TGTTTTAAAT 600

CAGGAGAATA GTAACGATTT TTTCCTTTTT TGACGAACTC TATTCCGTAA CGATCAATCA 660

ATTTAATCAT GTACCTAATA TTAGAATTGT TTATCCCAAA TTTATTTGAA AGCTTCTCTA 720

AGCTATATCC TTGTTTTCTA AGTTCATAGA TCTGAACTTT ATCATCATAA GTTAGTTTCA 780

TAATAAAAAC ACCCCAAAAG TTAGATTTTT TCTGTCTAAC TTTTGGGGGG CAGTTCATTC 840

AACACCTGAT ACTATGCGTT TTTCTTATTT GAAATACTTT TTACTCAACC TCTTTATACT 900

CAATGAAAAT CAAAGTGCAA ACTAGAAAGC TAGCCTCAGG CTGCTCAAAA CAGTGTTTTG 960

AGGTTGCAGA TGGAAGCTGA CGTGGTTTGA AGAGATTTTC GAAGAGTATT ACTTAATCTT 1020

CTTGATACTT TGACTAAGAA TAAATCCTAC AATCATCCCT ACCATATTTT GCATAAAATT 1080

CGGTAGAATT TCTGGGAGGG CTGCTGCCCA GCCATTCATC AAAGCAGAAC CCAAGGCGTA 1140

GCCTCCTACC ATGGCAATAG TTGCTAAAAT AAGGCCTAAC CACTGACTTT TTCCTTTAAA 1200

TCCTGCGAAA AATCCCTGCA AGCCATGGTT GACCAAGCTA AAGAACATCC ACTGAGGGTA 1260

GCCTGATAAG AGGTCAATCA AGAAACTTGC TAGTCCTCCG ACTACCGCTC CTTCACGACT 1320

ACCAAAGTAA AAGGCCGCAA AGAAGACACC AGCATCTAAA AGAGTTAGAA TTCCTGTAGG 1380

TGTTGGGATT TTTAAGAAAT AACCTAGAAC CACAGAAAGG GCGGTTAATA GGGATACAAG 1440

GGCGATTTTA GTTGTTTTTG TTTGCTTCAT ATTGTCTTAC TCCATACTGA TCTGCTTGTG 1500

CAATAGCACG ATAAACGAAA GCCTTAGAGC TTTCTACTGC TGGCAAAAGT TTATCACCTT 1560

TAACCAGGTG ACTGGCAATG CTAGAGsCAA AGGTACAACs TGCACCAGCA TTTTGGCCTT 1620

GGATAACTGG ATTTTCTAGG ATAGTAAAGG TCTGTCCATC ATAAAAGACA TCCACAGCCT 1680

TGTCCTGACT AAGACGATTG CCTCCCTTGA TAATGACTGt GGCGCTCCTA AATCATGCAA 1740

TTTCTGCGCT GCAGTTTTCA TGTCTTCCAA GGTTTTAATT TCCTGACCGG ATAATAATTC 1800

TGCTTCTGGG AGATTAGGCG TAATCACACT GACATAAGGG AAAAAGCGAA TCAACTCTTG 1860

GCAGAGCTCA CTGACAGCTA CATCATGCGT TTCCTTGCAG ACCAAGACAG GATCCAACAC 1920

CACAGGTACT CCTGGGCGTT GTTTGATAAA GTCCAAGGCC TTCTCAGCCA CGCTGACAGT 1980

AGGGAGAAGA CCAATCTTAA TTCCCCCAAA TTCCACATCA CGCAAGCTAT CTAATTCATG 2040

TTGAAAAATG GTATCATCAG TTGGAAAGAC TTCAAATCCT TTTTCTGTCA AGGCTGTCAA 2100

ATAAGTCACT GCTACAAACC CATGCAAGCC GTTCAAGGTT TAGGTAGCCA AATCAGCTGA 2160

CAGTCCACCA CCACTAAAAA TATCATTTCC AGAAAGTGCT AAAATACGAT TATTCTTCAT 2220

AACGAATCTC CTTTAAATAC AAACCATTTG GTGCTGCAGT GGGACCTGCA AGTTGCCTGT 2280

CCTTCTTCTC CAAGATGAGA TCAATCTGCT CTACTGGCAT GCGGTTGTTA CCGATTTTGA 2340

GAAGAGTCCC CACCATATTG CGAATCTGTT TATACAAGAA ACCATTTCCT GAAAAGGTAA 2400

AGGTCAAAAA TTGTCCTGTC TCATCGACTA TTAAACTAGC TTCTGTGATG GTGCGAACCT 2460

TATCCTCTAC ACTAGTCCCA GAGGCTGTAA AACCGGTAAA ATCATGGGTT CCCTCTAGCT 2520

TTTTGATTGC AATCTGCATT CGTTCCACAT CGAGTGGGTA GGGAAAGTGG GTGGCATAGT 2580

GACGGCGCAT CGGATTTTTG GGACGTCCTC TATCCACAGT AAACTCATAG GTCTTGCTAT 2640

GCTTGGCATA ACGGCAATGA AAATCATCTG CCACAAGCTC AATCGAAATC ACATCAATAT 2700

CTTCAGGAGA CTGGGTATCC AAGGCAAAAC GGAGTTTCTC CTCATCCATC TGATAAGGCA 2760

GGTCAAAATG AATCACCTGT CCCAGGGCAT GAACCCCACT ATCTGTCCTA CCAGCACCGT 2820

GAACAGTAAT GGCTTGCCCT TTATTTAATC TGGTCAAGGT TTTTTCAATT TCTTCCTGAA 2880

CGCTACGCGC ATGAGGCTGG CGCTGAAAGC CAGCAAAGGC ATAACCATCA TAGGAAATAG 2940

TTGCTTTATA TCTCGTCATA GCCTCTATTT TATCAAGAAA TTAGTCTGTA AACAAGGACC 3000

TAAAACAAAT ATTGTATGGG TATAAAAATC TCATACTCTT CGAAAATCTC TTCAAACCAC 3060

GTCAGTTTCC ATCTGCAACC TCAACACACT ATTTTGAGCA ACCTGCGGCT AGCTTTCTAT 3120

AGTAGATTGA AATAAGATAT GAACAACTCT ATTAGGAAAG TCAAATTAAT TTCTAGAAAT 3180

ATTTTAGCAG CTACAGCGTA CTATTCCAAA CTCAATCAAC TATAGTTTGC TCTTTGATTT 3240

TCATTGAGTA TCAAAAGAAA AACTTAGGAA TCAATCCTAA GCTCTCTTCT GAAGTAGGTA 3300

CATGACAAAG ATAGAGATTA CAATCAACCA ACCTCCTAAG ATACTAAAGA CCAACATCCC 3360

ATTGTGAGTT AGTAAGCCAA TTGCACCTAG AACGAATGGG GTCGTAAAGG CTCCGAAACT 3420

ACAGCCTAAT ACAGCAAATG AAGTTGCTTG ATTGAGGAGT TTAGCTGGAA TTCGTTCAGA 3480

GACAAGTTGA AAGACCGTCG TCAAGACTAC ACTATAGGCA AATCCAGCCA GAACACTTCC 3540

TGCTACTACC ACCCACAAGG ATGAAGACAA GGCAATCACG ATTTGCCCCA AGCCAAAGGT 3600

AATACCAGAC CAGAGGAGCA GTTTCTCTTT AAAGATAGAA ATCAAGAAAG AAAAACTCAC 3660

CCCAGCCACA ATCCCGATCA ACTGCATGAT ACTAAGAACA AAACTAGATA ACTGGGCATC 3720

CCCCAATCCT CTTTCCACCA TCAAACTTGG AATACGGATG GTAATAGCTG TATTGGTACA 3780

AACTACAACT GCCGCTTCGA TAGCTAAGGT AAAAATCAAG CCTTTCATTT CTCGAGTTAA 3840

ACGACTTGCT TCCTTCGCTC TTTTCTTGAC TTCTTTCTTT GATTTTCCAT AAGGGACAAA 3900

GAGCAGATAA AGGGGCAGCA CCAAAAATCC AGCACTATAG GCTAGAAAGA TAGCTGTCCA 3960

ACCAAAGGCC AACAACTGAC CGACGGCCAA GGTAATGAGA GAAGCTCCAA CGACCTCTGC 4020

AGAAGCGCGT AGCCCTAACA TCTGAATTCG CCTTTTTCCT TGGTAGCGTT CACTGATAAT 4080

AGAAATGGCC TTGGCATTGA TCATCCCAAG ACCCAAACCA AAGAGAAGCC GTGTTCCAAA 4140

GACAAAGGGA TAGGCTTGGT ACCAGAAGGG AGCTGTACCG CTCAATGATA AAATCAGCAA 4200

GCCCAAACTA ATCTGTAAGC GCTCAGGAAA TATTTTTTCT AAGAAACCAT TTAGCAGTAA 4260

CATCATCATG ATTCCAAAGG AAGGCAAGCT CACCAAGAGC TCAATTTGTT CCTTAGAATA 4320

ACCCTGATAA TAGTCAAACA TGGCTGGTAG GGCACTCGAA ATGGAAAAGG AGGTAATCAA 4380

AACGAGGGAG AGAGCCAAAA TGCTGGCCCG TTCTAAAAAT TGTTTCATGA AATCTCTTTC 4440

TATATTTCTC TTAATCTTCT ACTTTTTTGA TAGTTATCAA ATAAGCAAGA AAAGAAGAAG 4500

CCTCATTGGT TTGTAGACTC CTTCTTAAAT TCGAAAATGA ATCCCTTGTA TCTTATACTC 4560

AATGAAAATC AAAGAGCAAA CTAGGAAGCT AGCCGCAGGT TGTTCAAAAC AGTGTTTTGA 4620

GGTTGCAGAT GGAAACTGAC GTGGTTTGAA GAGATTTTCG AAGAGTATTA GGATGACTTT 4680

CTCTTGATTT GCTTGATAAA GTAGAAAATA AATCCTGCTA CCATATAGGC AACAAAGATA 4740

ATCAGACACC ACTTAAACAC AACATTCCAA CCCTTGTTCA CATTCAAAAA GAAGTAAGGG 4800

AAAGGATTAT CCTTGGCATT TGGAATATTG AGTTTTAGAA CCAAGCCATT AAAAAGAGCA 4860

AACATCATAT ACAGAAAGGG TAAAATGGTC CACACTGCTG GATCCCAAAT CTTGTATTGA 4920

CCCTGTTTGT CAAAAAAGAG GGTATCCGCT AAAAACCAGA TGGGAACGAT ATAGTGGCAA 4980

AGGAAATTTT CTAGGGTATA GAAATTAGTC GCAATGGGCG CCAAGAGGAA ATGGTAAATC 5040

ACACAGGTAA TCATGATACT CATGGTGACC CCACCTTTTA AGCGCAAGAG ACTTGGCCTT 5100

TGCCAATTTT CACCTACACG GCTCATAACC TTTAGAAGAT AAAGGGTAAA AATAGTTACC 5160

AAGAGGTTGG ACAGAACCGT GTAATAGAGA AGCATCCCAA AACCACCATG CTTAGTAATT 5220

TCAAGATAAA CTCCCGTAAA AGCCGCTAGA AACAAGAAGA TACGGCTATA AAATACAAGT 5280

TTATAGTGTT TTGACATGCT TAAATCTTCC TCACAAACTC TGATTTAAGT TTCATGGCAC 5340

CAAAACCATC AATCTTACAG TCGATATTGT GGTCGCCTTC TACGATGCGG ATATTTTTCA 5400

CGCGCGTCCC TTGTTTCAAA TCTTTTGGCG CACCTTTTAC TTTCAAGTCC TTGATGAGAG 5460

TTACTGTATC ACCATCAGCC AATTTATTTC CGTTGGCATC GATAGCGACA AGACCTTCTT 5520

CTACTTCTGC AACTTCAGCA GGATTCCACT CATGAGCACA CTCTGGGCAA ACCAGTAGGG 5580

CACCGTCTTC GTAGACATAC TCTGAGTTAC ATTTTGGACA ATTTGGTAAA TTGTTCATGG 5640

TTTCTCCTTA TCATCATTCA CTATTCTTTG AAAATCAAAA TTTCTCGAAC AGCAACTATT 5700

ATACCCTAAA ATCAGCATTT TGACAAATTT AGAAAAAAAC CGATATCAAT CTATCGGCTT 5760

TTCTACATTT ACATTCTTTT TTCAGCTTCT GCTTTGATTT TTTCAACTAC TTCTTGAATG 5820

TTCAAACCAG TTGTATCAAG GTAGACAGCA TCCTCTGCTT GTTTGAGAGG AGAAGTCTCA 5880

CGATGACTAT CCTTGTAGTC ACGCGCAGCA ATTTCCTTTT TTAGGGTTTC AAGGTCTGTT 5940

TCAATTCCCT TGGCAATATT TTCCTTGTAA CGACGCTCTG CTCTCTCATC AACAGAAGCT 6000

ACTAGGAAAA TTTTCAATTC TGCTTGTGGC AATACAACAG TTCCAATATC GCGACCATCC 6060

ATGACAATCC CGCCTTGCTG GGCAATTTCT TGTTGGAGAG AAACCAGTTT CTCACGCACT 6120

TGAGGAATTG CTGCAATAGC AGAAACATGA TTGGTCACTT CATTTTCACG GATAGGATGG 6180

GTAATATCCA CATCTCCTAC AAAAACAAGC TGGTCTCCAG TTTCTGAACG TCCAAAGCTG 6240

ATTGGATGCT GGTCCAACAA GGCTAGAAGG GCTTCGACTT CTTCAACTCC TAATTGGTTC 6300

TTAAGAGCCA TATAGGTCGC TGCACGATAC ATAGCTCCTG TATCAAGGTA GGTGAATCCA 6360

AAATCCTTAG CAATAATCTT TGCGACCGTA CTCTTACCGC TGGAAGCAGG ACCATCAATA 6420

GCAATTTGAA TTGTTTTCAT ATCGGCTCCT ATTTTATTTT TATAACATCA CCTGGATTAG 6480

CAAACCAAGA TCCTGTAGCC ATGTGCCCAG GATTCAAGGC CTCTAACTGA GCAATGGAGA 6540

TTCCTGCACG AGCGGCAATA GCTGCTTCCC CTTCTCCTGC GAGAACTTTA ATCGTTCCTT 6600

CAGGATTAGC AGCTTCTTCT GAACTACTAG AAGTAGATTC TGGCTCTGAA CTCTGCTCAG 6660

GCTGAGAACT ACTTGAAGAT GAGATTTGTA CTACACTGGC ATCAGAATCA TGAAAGCCTT 6720

TTAAGGCTGC TGTGCGATTA CTCCCCCCCG ATGATAGATA GATGAGAACG ATGACCATCA 6780

CCACCACAAT TACAAAGAAA ATACTAGCTA GGATCGTCAA AATACGATTA GCCATCCTAT 6840

CAGCCCCTCC GTGGTTTCGA TGCCGACGCT CTGCTCTTGA TTCTTCTTGA TCATAGATAT 6900

CTTCTTGCCA CGGTTCTTTT GCCATACCTT ACTCCTTGTT TTTTTTTACT TTTCTTATTA 6960

CAATATAAAT ATGAACATGA AAATCACACT TATACCTGAA CGATGTATCG CCTGTGGGCT 7020

TTGCCAAACT TATTCTGATT TATTTGATTA CCACGATAAT GGAATCGTGC GTTTTTACGA 7080

TGACCCTGAC CAACTGGAAA AAGAAATTTC TCCTAGTCAG GATATCTTAG AGGCTGTTAA 7140

AAATTGCCCA ACTCGCGCCC TGATTGGAAA CCAGGAAGCC TAAATCAATG GCGATAATCC 7200

ACTCCCTCTA GTTTAGCACA TTTCCATGTA AAATTATAGT CTTTTCACTT TATTTTTTTC 7260

TGTAAAATCA GGAAGGTCAC TTTTTTCTTT GATAAGATAA AGTGGTCTTT TTTTAGTCTC 7320

TAAATAAATC TTACTGATAT ACTTGCCGAG AATCCCAATG GTCAAGAGTT GAATGCCTCC 7380

AAGAAAGAGA ATAACAGCCA TCAGAGAGGT CCAACCAGAT GTCGGATTGC CCAAAATGAG 7440

GGTCCGAACC ACAACAAAAA AGGTCATCAG CAGAGAAAGA AAACAAGATA GGAGACCAGC 7500

TACAAAGGCT ATAATCAAGG GAAAATCTGA AAAATTAATA ATCCCTTCAA TGGAGTAGAA 7560

AAAGAGTTGC CTAAAACTCC AACTTGTCTT GCCAGCCTGC CTTTCGACAT TTGGATAGTC 7620

CAAATAGTAG GTTTTGAAAC CCACCCAGGC GAAGAGCCCC TTTGAAAAAC GATTGGACTC 7680

GGTCAAGCTT AAAATGGCAT CGACTACAGA CCTTCTCATC ATACGAAAAT CACGGACACC 7740

CGACGGCAGA GCTACTGGGC TGATTTTTTG CATGAGGCGA TAAAAGAGAA CAGCACAGAA 7800

ACTGCGAAAG AAGGGTTCTC CCTCCCGACT AGTTCTCCGT GTCCCAACGC AGTCCAAGTC 7860

TACATTTTTG TCTAATACAT TTTTCATCTC AAACAACATA CTAGGAGGAT CTTGGAGGTC 7920

TGCATCCATC ACCACCACCA AATCTCCTGT CGCATATTGC AAGCCTGCAT AAAGGGCTGC 7980

TTCTTTGCCA AAATTTCGAG AGAAAGAAAT ATAATGGACT GCCGGATTTT GCTCCCGATA 8040

GGCCTTTAAG AGTTCCAAGG TCCCATCACT TGATCCATCA TCGACAAAGA CATACTCGAT 8100

TTCTGTTTCC AAATCTGGAA GTAAAGCTTC CAGAGCCTGA TAAAAAAGAG GAAGTACTTC 8160

CTCTTCGTTT AAACAAGGGA CGATGATTGA AATCATCATC TTAGTCTTCA AATCCATTTG 8220

GATGCTTGCT TTGCCAACGC CATGCGTCTT CACACATTTG GGTGATGTCG AGTTCTGCTT 8280

CCCAACCGAG TTCTGCTTTA GCTTTTGCCG GGTCTGAGTA GCAGGCAGCG ATATCACCTG 8340

GGCGACGTTC TACGATGCGG TAAGGAATAG GACGGCCCAC CGCTTTTTCC ATGTTTTGGA 8400

TAATTTCAAG AACTGAGTAA CCTTTACCAG TTCCAAGGTT ATAAACGTTT AGTCCTGAAC 8460

CTTTTTGGAT TTTTTTCAAA GCTGCAACGT GACCCTTAGC CAAATCGACA ACGTGGATAT 8520

AGTCACGAAC ACCTGTTCCA TCTTCCGTAT CGTAATCGTC TCCAAACACT TGCACTTGCT 8580

CTAATTTTCC AACGGCTACT TGAGTCACAT ATGGCAAGAG ATTGTTTGGA ATACCGTTTG 8640

GATTTTCTCC CAAATCACCA CTCTCATGGG CTCCGATTGG GTTAAAGTAA CGAAGCAAGA 8700

CAACATTCCA TTCTGAGTCT GCTTTGTAAA TATCAGTCAA AATTTCCTCT AGCATGAGCT 8760

TAGTACGACC GTATGGGTTG GTCACTGAAA GTGGGAAATC TTCCAAGATG GGCACTGTGT 8820

GCGGATCCCC GTAAACTGTC GCAGAAGAAC TGAAGATGAT GTTTTTACAG TTGTTTTCTT 8880

CCATGGCTTT CAAAAGGCTG ACAGTTCCAG CGATATTGTT GTCATAGTAG GCAAGAGGGA 8940

TACGTGTTGA TTCGCCAACA GCCTTCAAAC CAGCAAAGTG AATGACACCA GTCGGTTCTT 9000

CCTGCTTGAA AATATCTCTG AGGGTATCTG TGTCACGAAT ATCTGCCTCA TAGAAAGGAA 9060

TCTCAACTCC TGTGATTCCT TCAACAACTT CTAAACTCTT ACGATTGCTA TTGACAAGAT 9120

TATCCACCAC AACAACTTGA TGACCTGCTT GGATCAATTC AATAACAGTG TGGGTTCCAA 9180

TAAAACCGGC ACCACCAGTT ACCAAAATCT TTTCTTGCAT CTTTTTTCCT CGATTCTCAG 9240

ATTATTTTTT CTTATTTTAC CATTTTTGAC AGGGAATGTC ATTTGCCATC CTAAACTACC 9300

TGATAAAATT TCAGTAAAAT GCTTATACTC TTCGAAAATC CAATTCAAAC TACGTCAACG 9360

TCGCCTTGCC ATGGGTATGG TTACTGACTT CGTCAGTTCT ATCCACAACC TCAAAACAGT 9420

GTTTTGAGCT GACTTCGTCA GTTCTATCCA CAACCTCAAA GCAGTGCTTT GAGTAACCCG 9480

CGGCTAGTTT CCTAGTTTGT TCTTTGATTT TTATTGAGTA TTATTCGCTT TTTACTCGTT 9540

TGACATAGTT TTCAATTGGG TAATTTAGAG GGTCCAAGGT CAACTCCTTG TCTTGGATCA 9600

GTTGGGCTAG ATGGTAACCA ATGATAGGAC CAGTTGTGAG GCCTGATGAA CCTAGTCCAC 9660

TGGCTGCATA GACACCAGTT AAGTCAGGCA CCTGCCCAAA GAAAGGAGAG AAATCACTGG 9720

TGTAGGCACG GATTCCAACA CGCTCAGATT TTGAAGTAGC TTCAGCCAAA ATCAGATAGT 9780

GAGTCAAGGT GGCCTCCTCC ATTTGTTGGA GCAAGGTTTC ATCTACCGTC AAATCAAATC 9840

CCATGTCATT TTCGTGGGTA GCGCCTAAGG ATAATTTCCC ACCTGCAAAG GGAATCAAAT 9900

CCCACTCCCC TTCTGGCATG ACAACAGGGT AATCTTCCAT GTCTTGGGCA AGCTGATAAT 9960

CTCGTAGTTG TCCTTTTTGA GGACGGACAT CCACTTCATA ACCTAAAGGC TCTAACATGT 10020

CCCCCAACCA AGCTCCCGTC GCCAAAATAA CCTGCTCAAA CTCCTCTTCA CCAATCTGGT 10080

AGCCTGATGC TAACGGTGTC AGAGTCACTT TTTCTTTGAC CAGCTTGACA TGACTGACTT 10140

CCAGCAAACG AGTCACTAAA AGTTGGCCAT CTACTCTCGC TCCACCAGAA GCATAGAGCA 10200

GGCGGTCAAA TCCCTGCAAA CCAGGGAATA ATTCATTAGC TGAGGCTTGG TTCAGAATGG 10260

CTAATTGCCC TATCAAGGGA GATTCTTCTC TGCGCTGGAG GGCCAGTTGA TAAAGTTCTT 10320

CCAAATTGGA TTCATCCTTT TTCAAGAGAA AGACTCCCGA ACGCTGGTAA AAGTCGATTT 10380

CTTGTCCTGA TTTCTCTAAA TCAGCTAATA AATCCACATA AAAATCAGCC CCCAAGCGCG 10440

CCATCTTGTA CCAGGCTTTA TTACGGCGTT TGGAAAACCA AGGACTGATA ATTCCTGCTG 10500

CGGCCTTGGT GGCTTGACCT TGCTCATGGT CAAAAACGGT CACCTCTAGG TCACTTTCTC 10560

TCGAGAGGTA GTAGGCAGCT GTTGCTCCCA CAATTCCTGC TCCAATAATG GCAACTTTTT 10620

TCATTGTCTT CACTTTCTAA CTAGATATGA TGGAAAGGAT TGGTTGATGC CTGACTAGGC 10680

AAGATATCAA TAGACCACCC CTTATCTTCC TTCCATTGAC TAAGAAGTGC TGCGATTTTT 10740

TCTACAAAAA TCACTTCGAT ATAGTGACCT GGGTCCAATG CAAGCAACCC ATCAGATAGC 10800

ATATCCTGAG CAGTATGGTA GTAGATATCA CCAGTGATAT AGACATCTGC CCCCTTTGCC 10860

AAAGCATCCT TATAGAAAGA CTGCCCGCTT CCACCACAAA TTGCTACTCT TGAAATAGGC 10920

TTCTGCAAAT CATCCTCTTG ATAATGCACC ATTCGAAGGC TATCTAGGTC AAAGACTTGC 10980

TTGACCTGTT GGGCCAATTC CCAAAATGTC TGAGGCTGAA TATTCCCAAT ACGTCCAATT 11040

CCACGTTCTG GACCTGTTTC CTGCAGATAA GTCGTCTCCT CGATTCCTAG CATCTGACAA 11100

AACCAGTCAT TGAGCCCATT TTCAACGATA TCAATATTGG TATGGCTGAC ATAAACTGCC 11160

ATATCATGCT TAATCAGGTC GATGTAAATC TGATTTTGCG GACGGCTGGC AAGCAAGTCC 11220

TTGATAGGAC GAAAGATAGG CGCGTGCTTG ACGATAATCA AGTCCACACC CTTTTCAATG 11280

GCCTCTGCCA CTGTCTCTTC ACGAATATCG AGGGCAACCA TGACCCTTTG GATACCCTTG 11340

TCTAAAGTGC CAATTTGCAG ACCACGGCTG TCTCCCTCCA TAGAAAATTC CTGAGGGCAA 11400

AAGGCTTCAT AAGCTTGGAT CACTTCACTT GCTAACATGG AGCACCTCCT TGATAGCTTG 11460

AATCTTATCT ACTAGAACTT GACGTTCTTC CAGATTTTTT TCTGGGATTT GTCCGAGGGC 11520

GAACTCTAGC TTCTCAGCTT CTTTTTGCCA TTTTTGGACA AATACTGGAC TGACTTCTTT 11580

GGACAAGAAG GGACCAAAGC GAACATCACT GGCTGATAGC TTCATTTGTC CTGCTTCCAC 11640

CACCAAAATC TCATAAAACT TTCCAGCTTC TTCTAAGATG CTTTCTGCTA CAATCTGGAA 11700

TCCATGATCC TGTAGCCAGA TACGCAAGTC GTCTTCACGA TTATTGGGCT GGAGGATCAA 11760

ACGCTCTACA TTAGCTAACT TCCCCAAACC TTCTTCTAAA ATCCTAGCAA TCAAACGACC 11820

ACCCATGCCA GCAATGGTAA TGACAGACAC TTGGTCAGTC TCTTCAAAAG CTGCCAAGCC 11880

ATTGGCTAAA CGGACTTGGA TTTTCTCCTT TAGGCCGTGA GCCTCAACAT TTTTAACCGC 11940

AGACTGATAG GGACCTTCCA CCACCTCACC TGCAATAGCG CTTTTGATTT GGCCTCTCTC 12000

AACCAACTCG ATAGGCAGAT AAGCATGGTC ACTTCCCACA TCTAGTAAAA TAGCCCCCTG 12060

TGACACAAAG GAAGCTACCA ATTCTAATCT CTTTGAAATC ATCTTCTCTC ACTTTCCAAA 12120

ACTCTATTAC CTCTTATTAT ACCACATTTC AATCTTCAAC TTCCCAGTAA TATAAGCACC 12180

TCTGGCGAAA GAAGTTTCAA TGTCCTAAAG TAATAAGTGA ATCCAATTGA AAGATTTTAA 12240

ACAATTTGCA AAAATGTCAA AAAATAAAAA ATAAACAGTT TATTCAGAAA ATTCTTGACA 12300

TATAAAAACA CATGGTAGAA TATAATTAGA AAGTTAGAAA AAATAAAAGT TTGACTAAAA 12360

TTTGTATTTG AAGGTGGTGT TCAGATAAGA AATTTAGTCA GACGAACCAC GAATTTGCTC 12420

TATGCTTTCT GGAATTTATC ATAACAGGAG GATACAGTCA TGGAACAAAC ATTGTTTGAA 12480

TTAGAACTAC TTCCAGAGGA AGATATCATT GTCACAGGTC TCCCTAAGTA TTGTTCTTTT 12540

ACTTGTTTAA TTACAGGTCG CTAGTTATAT TTTATATAAA ATAAGTAGCT TTACTTACGG 12600

AATAGGCTAG TGCTGTGTCT CTAGCCTATT TTAATAATTA GGAGTTTGTT ATGGATTTAT 12660

TAGAGAAAGA ATGTTTAAAA TGTGATAAAA ATTTCCAACA GGGTGATATT TGGAATTACT 12720

ATTATTTATC AGATAAGATG CCTGCACAAG GGTGGAAAAT ACACATAAGC TCCCAAATAA 12780

AAGACGCTGT AAATATTTTT AAGATTGTGT ATAAACTATC CCAACTAAAT AATTGTAGCT 12840

TTAAAGTTGT TAAAAATTTA GAGGAATTAA AAAAAATTAA TTCCCCTAGG GAAATGAGCC 12900

CTACTGCTAA CAAATTTATA ACTCTATATC CTAAGTCAGA ATCTGAAGCT AAGAGTATGA 12960

TTTGTAATCT TACGAATAGA CTGTCAGAAT TTAAGGCTCC AAAAATACTA TCTGACTATC 13020

AATGTGGAAT GCATTCTCCA GTTCATTATA GATATGGGGC TTTTTTAAAA AAACAAGCTT 13080

ATGATGAAAA AAATAAAAAA GTCATCTATT TATTGCTAGA TGAAAAAAGG AAGAACTATG 13140

TAGAAGATAA GAGACAAAAT TTCCCTAGTC TTCCTAGCTG GAAAATGGAT TTATTTTCAG 13200

AAGAAG 13206

(2) INFORMATION FOR SEQ ID NO: 34:

340

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13104 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

CCGGATCCAG CGAAAAATAT GCTCTTTGAT GCTGTAAGTG GTCAAAAAGA TGCTAAAACA 60

GCTGCTAACG ATGCTGTAAC ATTGATCAAA GAAACAATCA AACAAAAATT TGGTGAATAA 120

AAAATTTGTT CAAGGGGGGT GGAAATCAAA TCCCCCTTTG AATTTATCAA TAGAGACACA 180

AATAATTTAG CTTTCTTATA AAAAAGTAGT ATCCTATGAA AGGAGTTAAT ATGGAAAAGC 240

AACAACCTAG TAAAGCAGCC CTGCTGTCTA TCATTCCTGG GTTAGGACAG ATTTACAATA 300

AACAAAAAGC CAAAGGTTTT ATCTTCCTTG GTGTAACCAT CGTATTTGTC CTTTACTTCC 360

TAGCACTTGC AACCCCTGAA TTGAGCAACC TCATCACTCT TGGTGACAAA CCAGGTCGTG 420

ATAATTCCCT CTTTATGCTG ATTCGTGGTG CCTTCCATCT AATCTTTGTA ATCGTTTATG 480

TACTCTTTTA TTTCTCAAAT ATCAAAGATG CACATACGAT TGCAAAACGC ATTAACAATG 540

GAATTCCAGT TCCACGCACA CTCAAAGACA TGATCAAAGG GATTTATGAA AATGGCTTCC 600

CTTACCTCTT GATCATTCCA TCTTATGTTG CCATGACCTT CGCGATTATC TTCCCAGTTA 660

TCGTAACCTT GATGATCGCC TTTACCAACT ACGACTTCCA ACACTTGCCA CCAAACAAGT 720

TGTTGGACTG GGTTGGTTTG ACCAACTTTA CAAACATTTG GAGCTTGAGT ACCTTCCGTT 780

CTGCCTTTGG TTCTGTTCTT TCTTGGACTA TCATTTGGGC TTTGGCAGCT TCTACTTTAC 840

AAATCGTAAT TGGTATCTTC ACAGCTATCA TTGCCAACCA ACCATTTATC AAAGGAAAAC 900

GTATCTTTGG TGTTATTTTC CTTCTTCCTT GGGCTGTCCC AGCCTTCATC ACTATCTTGA 960

CATTCTCAAA CATGTTTAAC GATAGTGTCG GTGCTATCAA CACTCAAGTA TTGCCAATCT 1020

TGGCTAAATT CCTTCCTTTC CTTGATGGAG CTCTTATTCC TTGGAAAACA GACCCAACTT 1080

GGACTAAGAT TGCCTTGATT ATGATGCAAG GTTGGCTCGG ATTCCCATAC ATCTACGTTC 1140

TGACCTTGGG TATCTTGCAA TCTATTCCTA ACGACCTTTA CGAAGCAGCT TATATTGACG 1200

GTGCCAACGC TTGGCAAAAA TTCCGCAACA TCACTTTCCC AATGATTTTG GCTGTTGCGG 1260

CACCTACTTT GATTAGCCAA TACACCTTCA ACTTTAACAA CTTCTCTATC ATGTACCTCT 1320

TCAATGGTGG AGGACCTGGT AGTGTCGGAG GTGGAGCTGG TTCAACCGAT ATCTTGATCT 1380

CATGGATCTA CCGTTTGACA ACAGGTACAT CTCCTCAATA CTCAATGGCG GCAGCTGTTA 1440

CCTTGATTAT CTCTATCATT GTCATCTCAA TCTCTATGAT CGCATTCAAG AAACTACACG 1500

CATTTGATAT GGAGGACGTC TAAGATGAAT AACTCAATTA AACTCAAACG TAGACTGACT 1560

CAAAGCCTTA CTTACCTTTA CCTGATTGGT CTATCAATTG TAATTATCTA TCCACTGTTG 1620

ATTACCATTA TGTCAGCCTT TAAAGCAGGT AACGTCTCAG CCTTTAAACT AGATACTAAT 1680

ATCGACCTCA ATTTTGATAA CTTTAAAGGC CTCTTCACTG AAACCTTGTA CGGTACTTGG 1740

TACCTCAACA CTTTGATTAT CGCCTTAATT ACCATGGCTG TTCAAACAAG TATCATCGTA 1800

CTTGCTGGTT ATGCTTACAG CCGTTACAAC TTCTTGGCTC GTAAACAAAG TTTGGTCTTC 1860

TTCTTGATCA TCCAAATGGT GCCAACTATG GCCGCTTTGA CAGCCTTCTT CGTTATGGCG 1920

CTTATGTTGA ACGCCCTTAA CCACAACTGG TTCCTCATCT TCCTCTACGT TGGTGGTGGT 1980

ATCCCGATGA ATGCTTGGCT CATGAAAGGC TACTTCGATA CAGTGCCAAT GTCTTTAGAC 2040

GAATCTGCAA AACTAGACGG TGCAGGACAC TTCCGCCGCT TCTGGCAAAT TGTTCTACCA 2100

CTTGTTCGCC CAATGGTTGC CGTACAAGCT CTCTGGGCCT TCATGGGACC TTTCGGGGAC 2160

TACATCCTCT CTAGTTTCTT GCTTCGTGAG AAAGAATACT TTACTGTTGC CGTAGGTCTC 2220

CAAACCTTCG TTAACAATGC GAAAAACTTG AAGATTGCCT ACTTCTCAGC AGGTGCTATC 2280

CTCATCGCCC TTCCAATCTG TATTCTCTTC TTCTTCCTAC AAAAGAACTT TGTTTCAGGA 2340

CTTACAAGTG GTGGCGACAA GGGATAATTT ATCCCCGCCA CCCTTTTTCA TTTTATACTC 2400

TTCGAAAATC TCTTCAAACC ACGTCAGCTT TATCTCCAAC CTCAAAGTTG TGCTTTGAGC 2460

AACCTGTGGC TAGTTTGCAC TTTGATTTTC ATTGATTATT AGCAATTGTC ACTGTAAATA 2520

ATATCCTTGT AGCAAGCAAT TTTTCTCCTA GACTTGAAAT AAAGCGCATT TCTCTATATA 2580

ATAATACTCA TATAGAAAAC ACCTTTTAGA AAGATACCTA TGCTTCCATA TCCATTTTCC 2640

TATTTTTCAA GTATTTGGGG GGTTCGTAAG CCCCTGTCCA AACGTTTCGA GCTCAACTGG 2700

TTTCAACTTC TCTTTACCAG TATCTTCCTT ATCAGCTTGT CTATGGTACC CATTGCTATC 2760

CAAAACAGCT CCCAGGAGAC CTATCCGCTA GAAACTTTTA TCGATAATGT CTATGAACCT 2820

CTGACAGATA AGGTTGTCCA GGATCTCTCT GAACATGCTA CAATTGTCGA TGGCACATTA 2880

ACTTATACTG GAACAGCTAG TCAAGCCCCT TCTGTTGTGA TTGGTCCAAG TCAAATCAAG 2940

GAATTACCTA AGGACTTGCA ACTGCATTTC GATACAAATG AGCTAGTCAT CAGCAAGGAA 3000

AGCAAGGAAC TGACCCGCAT CTCTTACCGA GCCATTCAGA CTGAGAGTTT CAAAAGCAAA 3060

GACAGCTTGA CCCAAGCAAT TTCTAAAGAC TGGTACCAAC AAAATCGTGT CTATATCAGC 3120

CTCTTCCTAG TTCTCGGTGC GAGCTTCCTC TTTGGTTTGA ATTTCTTTAT CGTCTCTCTT 3180

GGAGCTAGCT TTCTCCTTTA TATCACCAAA AGATCACGCC TCTTTTCATT TAATACCTTT 3240

AAAGAGTGCT ACCATTTTAT CTTGAACTGT TTAGGATTGC CGACTGCTAT TACACTTATT 3300

TTGGGATTAT TTGGCCAAAA TATGACAACC CTGATTACTG TACAAAATAT TCTTTTTGTT 3360

CTGTATCTGG TCACTATCTT TTATAAAACA CATTTCCGTG ATCCAAATTA CCATAAATAG 3420

GAGATTTTTA TGCCCGTTAC GATTAAAGAC GTGGCCAAGG CTGCTGGTGT TTCGCCTTCA 3480

ACCGTAACCC GTGTTATTCA AAATAAATCA ACCATTAGCG ACGAAACAAA AAAACGTGTT 3540

CGCAAAGCTA TGAAGGAACT CAACTACCAC CCAAACCTCA ACGCTCGTAG CTTGGTAAGC 3600

AGCTATACTC AGGTTATCGG ATTAGTTCTT CCTGATGACT CAGACGCCTT CTACCAGAAT 3660

CCTTTCTTTC CATCGGTTCT ACGTGGCATC TCTCAAGTCG CATCTGAAAA CCACTATGCC 3720

ATTCAGATAG CAACAGGGAA AGATGAGAAG GAGCGTCTCA ACGCTATTTC ACAAATGGTC 3780

TACGGCAAGC GTGTAGATGG GCTAATTTTT CTCTATGCCC AAGAAGAAGA CCCTCTCGTA 3840

AAACTCGTCG CAGAAGAACA GTTCCCCTTC CTTATCTTAG GTAAATCTCT ATCTCCTTTC 3900

ATCCCACTTG TCGACAACGA CAATGTTCAA GCTGGTTTTG ATGCGACTGA ATATTTCATC 3960

AAAAAAGGCT GCAAACGCAT TGCCTTTATC GGAGGAAGTA AAAAGCTCTT CGTGACCAAA 4020

GACCGTTTAA CAGGCTATGA ACAGGCGCTT AAACATTACA AACTTACCAC TGACAACAAT 4080

CGCATCTACT TTGCCGACGA GTTTCTGGAA GAAAAGGGCT ATAAATTTAG CAAGCGATTA 4140

TTCAAGCACG ATCCACAAAT TGATGCTATC ATCACAACCG ATAGCCTCCT AGCTGAAGGT 4200

GTTTGTAACT ATATTGCCAA ACACCAGCTG GATGTCCCTG TTCTCAGCTT TGACTCGGTT 4260

AATCCCAAGC TCAACTTGGC AGCCTATGTC GATATCAATA GTTTAGAGCT TGGTCGTGTT 4320

TCCCTTGAAA CTATTCTCCA GATTATTAAT GATAATAAAA ACAATAAACA AATTTGTTAC 4380

CGTCAATTGA TCGCCCACAA AATTATCGAA AAATAAGAGA CTGGGCAAAA AGTCGTTAAA 4440

AGCAAAAAAG CATACTATCA GGTATTGAAA AAACTTGATA CTATGCGTTT TATTGTGGGA 4500

AGATTTACTT CCTTTTCTAC TGAAATTGAG TCTTTTCCCA AGATCTTTTT ATACTCAATG 4560

AAAATCAAAG TGCAAACTAG GAAGCTAGCC GCAGGTTGCT CAAAACACTG TTTTGAGGTT 4620

GTAGATGAAA CTGACGAAGT CAGTAACCAT ACCTACGGCA AGGTGAAGCT GACGTGGTTT 4680

GAAGAGATTT TCGAAGAGTA TTAATCACTA ATTATCTATC TCAACAAATC TTCCTAGAAT 4740

ATGAACATTT TCCGAGACAG AGACAAAGGA GCTTGGATCC ACTTGTGTCA TAATCTGTTT 4800

AAATTCATTA AACTCTGCAC GTGTAATGAC AGTGATTAAA ACTGCCTTTC TCTCGTGATT 4860

ATAGGTTCCT TCTGCATCGT GGATCATGGT TGCTCCGCGG TGCAATTTTT TATGGATTTT 4920

TTCAATTACC TTCTCTGGAT GATTTGTCAC AATCATGGCC TGCATACGCT TTTGCTTAGT 4980

AAAGACTGCG TCTGTCACAC GGCTAGAGAC AAAGATGGTA ATCATAGAAT AAAGAGCGTA 5040

TTTCCAACCA AAGGTCAAAC CTGCTATCAG CATGATAGTT CCATTTACCA AGAAAGAAAT 5100

ACTACCGACA TTCTTACCCG TTTTCTTACG AATAGTCAGG CTGACGATAT CCGTCCCACC 5160

ACTGGAGATA TTGTTTCGAA GAGCAAAACC AATCCCCAAA CCCATAACAA CACCCCCAAA 5220

AAGGGAATTG ATAATGGGAT CCTCTGTCAA GGTTGCCACA GGGACAAACT GGATAAAGAA 5280

GGAACTCATA GATACCGTGA TAAAGGTAAA GACGGTGAAC TTATGGCCAA TCTGATACCA 5340

AGCTAAGACC ATCAAAGGGA AGTTAATGGC GTAGAAGCTT AGCGAAATCG GAATATGAAA 5400

ACCAAACCAG TGATTACTCA AGGCAGAGAT AATCTGTGCC AGACCTGTTG CACCACTCGA 5460

ATACACATGC CCTGGTTGGA AAAAGAAATT AACTGCTACT GCTGATAAAA AACCATAGAC 5520

CAGAGAGGCC GAAATCTTCT CATCATACTT TTCTCGAGAG ATACTTTGTA AGACACGTAA 5580

AATTTTTATC TGATAAGCAA AGCGGCGCAG ATAATAGCGC CACCGCTTAA TTCGTTTTGT 5640

TTGTTTCATC TTCTTCTACT TGTAAGCTGA GTTCCTCTAG TTGTTTGAGA GCGACTGTTG 5700

ATGGAGCTTG TGTCATTGGG TCAGTTGCCT TGTTGTTCTT AGGAAAGGCA ATGACTTCAC 5760

GGATATTTTC TTCTCCAGCA AGCAACATGA CAAAACGGTC AAGCCCGATA GCCAAACCAC 5820

CGTGTGGTGG GAAACCATAG TCCATGGCTT CAAGAAGGAA ACCAAACTGG TCATTGGCTT 5880

CTTCAGTTGA GAAACCAAGA GCCTTGAACA TGCGTTCTTG AAGGTCTTTT TGGTTGATAC 5940

GAAGGCTACC ACCACCAAGC TCATAACCGT TCAAGACGAT ATCGTAAGCA ATGGCACGAA 6000

CCTTAGCCAA ATCACCTTCT AATTCATGAG CAGTCTCTTC CTGTGGAAGT GTGAAAGGAT 6060

GGTGGGCGCT CATGTAGCGG CCTTCTTCTT CAGACCATTC AAACATCGGC CAGTCAACCA 6120

CCCAAAGGAA GTTGAACTTA TCATTATCAA TCAAGCCAAG CTCTTTAGCA ATACGTCCAC 6180

GAAGGGCACC CAGTGTTGCA TTAGCCACTT CAAGCGTATC CGCCACAAAG AGAACCAAGT 6240

CCTTATCTTC AAGAACAAGC GCTGTTGTCA ATTCTTCTTG GATACCAGTC AAGAACTTGG 6300

CAACTGGTCC GTTTAATTCT CCATCAACCA CCTTGACCCA AGCAAGACCT TTGGCACCAT 6360

ACTGTTTGGC TACTTCCGTC ATCTTGTCGA TGTCTTTACG TGAATAGTTG TCCGCAGCTC 6420

CTGTGACCAC AATCGCTTTT ACAGCAGGTG CTTCTGAAAA GACTTTAAAG TCTACACCTC 6480

GGACCACTTC TGTCAAGTCC TGAAGCAACA TGTCAAAACG AGTATCTGGC TTGTCAGAAC 6540

CGTAAAGAGC CATAGCATCA TCGTATTTCA TACGAGGGAA TGGTAGCGTT ACTTCGATGC 6600

CTTTTGTTTC CTTCATCACG CGCGCGATCA AGCTTTCTGT AATATCTTGG ATTTCTTGCT 6660

CAGTAAGGAA GGACGTTTCC AAGTCGACCT GAGTAAATTC AGGCTGGCGG TCTCCACGCA 6720

AGTCCTCGTC ACGGAAACAT TTAACGATTT GGTAGTAACG GTCAAAACCA GCATTCATCA 6780

AGAGCTGTTT CGTGATTTGT GGACTTTGAG GAAGAGCGTA AAAATGCCCC TTATTAACAC 6840

GAGACGGCAC TAAATAATCA CGCGCCCCTT CAGGCGTTGA CTTAGAAAGG AATGGTGTCT 6900

CCACGTCGAT AAACTCCAAC TCATCCAAGT AGTTGCGGAT AGAGTGGGTC ACCTTGGCAC 6960

GAAGTTTAAG ATTTTCCAAC ATTTCTGGAC GACGAAGGTC AAGGTAACGG TAACGCAAAC 7020

GTGTATCGTC ATTTGCCTCA ATGCCATCCT TAATCTCAAA TGGTGTTGTC TTAGCTGTGT 7080

TAAGCACAAT AAGAGCTGTC ACGTTTAACT CAACCGCACC AGTTGGCAAC TTATCATTGG 7140

CTTGTCACGC GCAGCGACCT GACCAGTCAC CTCAATAACA AATTCGCTAC GAAGGcTTTC 7200

AGCTGTTGCC ATAACCTCTG CAGATACTTT TTCAGGGTTG ATAACCAACT GCATGATTCC 7260

TTCACGGTCA CGAAGATCGA TAAAGATCAA ACCACCAAGG TCACGACGAC GGCCAACCCA 7320

TCCTTTCAAG GTTATTTCTT GTCCGATGTG TTCCTCACGA ACACGACCAG CATACATACT 7380

ACGTTTCATT ATTTCTCTCC TCTTTTATTC TGTTACTATT TTACCATAAA AGCGCAGCTC 7440

TTCATGAAAA TCATCAGAAA AGTTTGCCAG TCTTTAAAAG TCAGGTGAAA GCCCTAAAAA 7500

TTAGCGCTAA TACTCTTCGA AAATCTCTTC AAACCACGTC AGCGTCGCCT TACCGTATGT 7560

ATGGTTACTG ACTTCGTCAG TTTCATCTAC AACCTCAAAA CCATGTTTTG AGCTGACTTC 7620

GTCAGTTCTA TCCACAACCT CAAAACAGTG TTTTGAGCAA CCTGCGGCTA GCTTCCTAGT 7680

TTGCTCTTTG ATTTTCATTG AGTATAATAC AAAAATCCGA TGAACTTCAC CGGACTCTTT 7740

TATTTTGAAT TTTTGCCTGC TTTACGCTTT TCAGCGATTT CGGCTGCCTT TCGAGGCAAG 7800

ACAATTTCCG TTATGTAAGC CGTCCCAAAA CGCAGTACAC CTGCAATAGG AGCAAAGACA 7860

ACTGCTAGAT AGTTATAGAA GAAATCGCCT TTGAAGGCAT AAGCTAGCGC TCCAATGATG 7920

AAAAATAGAA CGACTGCCTG AATCAGTGCT AATAAAATTA CTCGTTTCAT GTGACCTCCT 7980

GACTCTATTA TAGCATGAGA ATCATCAAAA AGCCGACTAA ATTATTCAAA GCGTGAAGAG 8040

AAATACTGTA GACCAGACCT TTTCTGCTAA TGTAAGCCAA ACCCAAACTA AAACCAAGGC 8100

TAAAATAGAC AAAAAATTGT TGCACATCAC CTGGAAAATG AATCAAGGCA AATAGAAGAC 8160

TAGATACCAG AAGAAAAATC AGGGTTCGTT TACTATTGTC CTGCTTAGGA AAGAGATAGC 8220

GTGCTAACAT CCCTCTAAAA ACAATCTCTT CCGTCAAAGG AGCAAAAATA ACCACAGCAA 8280

AGAATGAGAA AAGTGGTTGA GACAAGGTCA AGTCTGTCGC TATTTGCTGA TTTACTGAAG 8340

GATCATCTGG CAAGAAGAAT TGAACGACCA GAGATAAGAA CCAAACCAAG ACAGGAAGCC 8400

AAATAAATCG ATTAAAGCCG CTCTTCTCAA TATGAACAGG AGCCTTCTGA TACCATTTGT 8460

AAATGCCGTA CACATATACT CCAGCCAAGG CCACATAGAG TAGAGTAACA GCATAGGGTG 8520

AAGCGCCTAA AGCAAGCGAC GCAGTCGCGA GCCCCTGAAT AAAGCCATAG ATAAATAAAA 8580

AGGATAGAAG GGCTAGAAGA ATCCAGCCAA GGTTTTTAAG TAATTTCATA GATAACTCCT 8640

TTATTTGAAA TAACGTTTTA CCATAGGTAA CTGCATCACA TTGATATAAA CATGGATGGC 8700

TCCTACAAGC AAGAAAGCTA GTAACTGAAT CTCTCCTGTC AAGAAAGAAA TGATAATAAG 8760

AAAAATATAT AAGGCTGGTA AGACATATTG GTGTAATTGG AATAAAATTC GAAAACTCTG 8820

TTCCAAATTA GCCTGACGCT CCCCTTCATC ATAAGAATTT ATATAGTTCA AGACATCCTT 8880

TGGTGTAGCG AAAAATTCCA AATCAAACTG ACGAACAATC GCAATGGTTT TAAAAAGAGA 8940

TTTTTGAGCG ACTAAGAATA CCACAAAGAG TAAGAAAGAA AGGAAAAATG TTTGAGGGTT 9000

TGTATGCAAT ATAATCACCT CACTTAATGA AATAAAAATA GCCAATGGAA TCGCTACACC 9060

TGTAATATTA AAAGCAATGG TTCCAAACTC AAGATTCCGA TACATTTGCA CATAATAGGT 9120

TTCATTCAGA TCGTCATCCA TTTCCTCTTG ATACAAAGAA TGAAATTTTC TGCTTTTCTT 9180

TAAGAAATTG AAAGTCAAAA ACATACTAAT GAAACCTATC AGTAAACAAA TAGCTGATAT 9240

CCATGGCATC AAGGCTTTTA CATCTAAAAT AATTTCGTGG GATTCGACAC GTGCCTTAAA 9300

CATCCCTACA AACATGCCCA AGAACCCCCC AAGACAATAG ACATCAAAAA TAACAATCTA 9360

CGTTTCTTTT TCATATTCAT TCTCCTTTTT CACTTGCTAG ATTTTTGGAT TTCTTTTCAA 9420

TCCATTCAAT TACTGGGATG AGAGCAAAGT AGACCCAAAC AAATTGGTCG CTTTGATAGG 9480

GATTAAACCA GCTTAGGTCC ATCCCAATCA GTAGAAATAC GCTGACTAAT AAAGCTATGA 9540

CCACTACATA ATAAATCACT TTATACTTGT TCATCACTCG TCCTCCTCCA AACGAAATAC 9600

CGATTCGACT GTTTCGTTGA AAATTTGAGA TATTTTCAGG GCAATGATAA TGGATGGGGT 9660

GTACTCATCC CGTTCTAGTA GGCTAATGGT CTGTCTGGAA ACCCCTGCCA GTTTGGCTAG 9720

GTCGGTTTGA TTGAGACCAT CGCGAGCTCG AAGCTCTTTT AGACGATTTT TTAGTTGCAT 9780

GTTACACACC TACTCTCCGT CAAATTCAAC GGTTTGGATA TCCTCAATAC GTTGCAACTT 9840

GAATTTTTCT TTTCCCGTAT TATCTACACG TCGTAGCTTT ACCCATTCCT CATCAACATC 9900

CACAACTTCC CAGTTATCTG GCCCAATATA CACTCCCGTT ATAATTGGTT CCTTTCCAAT 9960

CATTTCTTGT AATAATCTCG ACATTTCTGC GTTTCCTTTC TCTTTTCGCT CAAGTCTTTT 10020

GATTTTATTC TCTAGTTTCT TGATTTTTTT AGAATTATTA GAATAAAAGA AAATCATAAA 10080

TAGTATAAAT CCTAGTACCC ACATTATAAC TCCTTTCTGC TTCCTATTTC TTAACTTGAA 10140

TTCATTGTAA CATATCTTTT TCTTTTTGAC AAGTATAGTT GTCAAAAAAA TTATGATTTT 10200

TGTCATTTTG CAAAAGAAAA AGGTCAGGAG TAGGTTCCTG ACCACTTTAT CTATCATTAA 10260

TACTCTTCTA AAATCTCTTC AAACCACGTC AGCTTCACCT TGCCGTAGGT ATGGTTACTG 10320

ACTTCGTCAG TTTCATCTAC AACCTCAAAA CCATGTTTTG AGCTGACTTC GTCAGTTCTA 10380

TCCACAACCT CAAAACCATG TTTTGAGCTG ACTTCGTCAG TTCTATCCAC AACCTCAAAA 10440

CCATGTTTTG AGCTGACTTC GTCAGTTCTA TCCACAACCT CAAAACAGTG TTTTGAGCAA 10500

CCTGCGGCTA GCTTCCTAGT TTGCTCTTTG ATTTTTATTG AGTATAAAAT CCTAGTTTTT 10560

CAAAGATTTC TGAGAAGTTT TGGCTGATTG TCTCAAGTGA CACTTGCACT TCTTCTCGGG 10620

TTTGGTTGTT CTTGACCGTC ACTTGTCCGC TTTCGACTTC GCTCTCTCCT AGGGTGATGA 10680

GGGTCTTAGC CGCAAAGACA TCGGCTGACT TGAACTGAGC TTTTAGTTTA CGGTTGAGGT 10740

AATCACGCTC TGCTTTGAAA CCTTGTTGGC GAAGAGCCTG TACCAATTCC AAGGCCTTGA 10800

TATTTGCCCC TTCGCCCAAG ACTGCGATAT AGACATCTAG GGCGTTTTCG ATAGGGAGGG 10860

TCACACCTTG CTTTTCAAGG ATGAGAAGCA GGCGCTCTAC ACCAAGTCCA AAACCAAATC 10920

CAGCAGTTTC AGGGCCTCCA AAGTAAGCAA CCAAACCATC GTAGCGACCA CCCGCACAGA 10980

CGGTCAGGTC ATTGCCCTCA ATCTCTGTGA TAAACTCGAA AATGGTGTGG TTGTAGTAGT 11040

CCAGACCACG CACCATATTG GTATCGATGA TGTAATCTAC TCCAAGATTT TCCAACATCT 11100

GACGCACAGC ATCAAAATGA GCTTGGCTTT CTTCATCAAG AAAGTCCAAG ATAGACGGCG 11160

CATTCTCTAC TGCCACCTTG TCTTCTTTTT CCTTAGAGTC CAAGACACGA AGAGGATTTT 11220

CCTCCAAGCG ACGTTGGCTA TCCTTAGACA AGGTCTCCTT GAGCGGTGTC AAATAGTCAA 11280

TCAAGGCTTG GCGGTAGGCT GCACGGCTCT CAGGATTTCC AAGAGTGTTG AGGTGCAATT 11340

TGACACCTTG AATACCGATT TCCTTCAAAA AATGGGCTGC CATAGCGATT GTTTCCACAT 11400

CGGTAGCTGG ATTGCTAGAG CCAAAACACT CAACACCAAT CTGGTGGAAT TGGCGCAAGC 11460

GCCCTGCCTG TGGACGCTCA TAACGGAACA TAGGTCCCAT GTAGTAGAAC TTGCTTGGCT 11520

TTTGCACTTC TGGGGCGAAA AGTTTATTTT CCACATAGGA ACGGACAACG GGTGCAGTTC 11580

CTTCTGGACG GAGGGTAATA TGACGGTCAC CCTTGTCATA AAAATCGTAC ATTTCCTTGG 11640

TTACGATATC CGTTGTATCT CCGACAGAGC GACTGATAAC CTCGTAATGC TCAAAAATAG 11700

GCGTGCGCAC TTCTGCATAG TTGTAGCGTT TGAAAATCTC ACGGGCAAAG CCCTCAACGT 11760

ACTGCCACTT AGCAGACTCA GCAGGTAAAA TATCCTGCGT TCCTTTTGGT TTTTGTAATT 11820

TCATAGGGAA TCCTCTTTAA ACTTAATAGT CTTATTTTAC CATAAATAGA GGGATTAAAA 11880

CAGTAAGAAA AAAATTAGGA TTTAGATATC ATTTTTGAGA TTAAGAATTG TCAAAAAAAT 11940

AGCTAGCAAG GAAAGACCAA CAAATAGCAT CCAAGTCAAC TGTATATTCC ATACGGCTAC 12000

TAGTGAAAAA CAAGCTGTTC CCACAGGTAT GGATAAGGTA AACAATAGAC CTAAAAAATT 12060

ACTAGTACGA GCTAGAACCT CTGGAGCTAG ATTTTTCATG AGCATGGCAC TAATCTTTGG 12120

TTGAACTTTA CCAGACACAT ACAGAGTAAA GAAGAGAAAT AGCAAACCAA GCACGACTTG 12180

ATTGAATAAA TTAGCCAAAC CAACTAGACT AAGTCCTACG GTCTCCCACA TCATCAATCT 12240

AGGCAAGGAC TGCTTCCCAA AATAATCATT GCCCGTAAGG CTACTGATGA TGACTGATAC 12300

TAAAACACAG AATTGATTGA TAAATAGTGC CTCTGTATAA GAAAAATTCA AGAGAGAATG 12360

GCTCAAAAAG AAGATATTAT AAATTCCACC CAAAGCGCCA CCCAAGGAAT TAATAAGCAA 12420

GACAGCAAAG AGCATAAAAC CAAAGTTTTT CTGTCCACTT TTAAGAAAAA CGAGACGTAA 12480

ATTTCGGTAA ATTGTTAGGA ACTGGTCTTT GATAGAAAGC TTCTCATTTT TTAAGTTTTC 12540

ACCATCAGCA GATGACATTG ACAGGCTCAA TTTGCTTTTT CCTAAAAAGA GGATAGTGGC 12600

TGATACTAGG AAAAAGCAGG CATTGATTCC CGCAACGAGA GAAAAATTGT TGACCGATAG 12660

AGCTAAGAGC CAGACTCCGA AAGCTTGACC ACCAATAGCT GAAATATAGG TGATGAACTG 12720

TGAAAAAGAA TAAGCCTCCA TCAGATCATC TTCAGCTACT TTTTCCTTAA TAAGAGGCAT 12780

ACGCAGGCCA CCTGCAAAAT CACTGATGAT ATCACTAATG ACATTGATCA AACACAGGCT 12840

AGAAAAGGCA AAGAGACTAG CTTGCTGAAC AACTAGGGCT GCTAGAAAAA ATAGAACCGC 12900

CTGAAACAAA CCGCTATAGA CCATCCATTT GACCTTGTCC CTCGTGTAAT CTGCCCGAAT 12960

CCCTGCAAAA ACTGTAAAGA GGGTCGGAAG AATCATGACA ATATTCGCCA TAGCAACAGC 13020

AAAAGATGCT TGTGACAAGG TCGATGCATA GACGATAAAG ACCAGGTTGA AAATCGAAAC 13080

ACCAAAAGCA TTGAAGAAGC GTGG 13104

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19250 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CCGGGCAAAT AGTTTTGAAC TTTTCATCAT TTTCTCCTTT AAAACTTTCT CTCCATTATA 60

GACTCTTTTC AGAAAGTTGT CAACAGAATT TTCAGAATTT TTGAAAATTA TTTTTCAAAC 120

AACATCTTTG CAAAAAATAT GAATATCGTA AGCGCGTCAT AACAAGGTAT CTATCATTCA 180

TGGAGCTCCT CCTGTATACT ATTAGTAAAG TAAATATTGG AGGATATTTT AATGCCACAA 240

CCTATTGTTC CTGTAGAGAT TCCACAATCT CGTCGTTTTG ATTCTAAAAA GAGAAATGAT 300

ATTCTTCTTA AAATTCGTAT TGGCAAGCTT GAAGTAAGTT TTTTTCAATC TCTCAATCTC 360

GAAATGATAG AACAGCTTTT GGATAAGGTG TTGCTCTATG ACAATTCATC TATCTAGCCT 420

AGGGCAGGTC TATCTCGTGT GTGGGAAAAC TGATATGAGA CAAGGAATCG ATTCACTGGC 480

TTATCTCGTT AAAACCCACT TTGAATTGGA TCCTTTCTCC GGTCAAATCT TTCTCTTTTG 540

TGGTGGACGT AAAGACCGCT TTAAAGTCCT TTACTGGGAT GGTCAAGGAT TTTGGCTACT 600

ATATAAACGC TTTGAGAACG GCAGACTGAC TTGGCCCAGT ACAGAAAAGG ATGTCAAAGC 660

TCTCGCACCT GAACAAGTAG ATTGGCTGAT GAAAGGCTTT TCTATCACTC CAAAAATATA 720

GTAGATTGAA ACTAGAATAG TACACCTCTG CTTCTAAAAC ATTGTTAGAA ATCGATTTTA 780

CTGTCCTGAT CGATTTGTCC TGTTATTATT TCATTTTACT ATAAATCCAT CAGAAAGTCG 840

TGATTTCTAT TGAAATGAGG ACTTTCTTTT TATACTCATC TGCTTTCAAA AAGCACTCTA 900

GTCCATCTCC GATTAACGAT GGACTTTATC ACCTCCTTCT CCAGTCCTTG TATAACATCT 960

TGAAGTTGAT TCATGACATC TTCCAAAGTT CGAAAGGCTT TATTCTTAAA TCCACGTTTA 1020

CGAATCTCTT TCCACACTTG TTCAATGGGG TTCATCTCTG GTGTGTATGG AGGAATAAAT 1080

GCAAAGCCAA TATTAGTCGG AATCTTTAAG GTACTTGATT TATGCCATAT AGCATTGTCC 1140

ATAACGAGTA AAAGATAATC ATCTGGATAA GCTTGTGAAA GCTCCTATTC CTAAAGCCCC 1200

TTTATAACCT CTTGCGAGAG AGACTATTGA CTCAGCCCTT ACTTCATGCG GATGAAACCT 1260

CCTATCGGGT TCTAGAGAGT GATAGCCATC TGACCTACTA TTGGACTTTT TTGTCAGGTA 1320

AAGCAGAGAA ACAAGGGATT ACGCTTTACC ACCATGATCA GTGTCGAAGT GGTTCAGTAG 1380

TACAAGAATT CCTAGGAGAT TATTCTGGCT ATGTTCATTG TGATATGTTG CGGCAGTAAC 1440

TTAGGACTTT AGTCCTCTAG TTCTGCCTAT GCGATAGCAG TCCAAGGTTT AGGAGTAAGG 1500

CGACGCTAAG CTTGGTAAAC TGCGAACAGC TAGAAGCTTA TCGTCAACTG GAAGAAGCTG 1560

CACTTGTTGG ATGTTGGGCG CATGTGAGAA GGAAGTTTTT TGAAGTGCCC CCCAAGCAAG 1620

CAGATAAATC ATCCTTAGGA GCTAAAGGTT TAGCCTATTG TGATCAGTTA TTTTCCTTGG 1680

AAAGAGACTG GGAGGCTTTG CCAGCTGATG AACGGCTACA GAAACGTCAA GAACATCTCC 1740

AACCCCTACT GGAAGACTTC TTTGCTTGGT GCCGTCGTCA GTCAGTTTTA TCGGGTTCAA 1800

AACTAGGAAG GGCAATTGAA TACAGCCTCA AGTATGAAGA AACCTTTAAG ACCATTTTAA 1860

AAGACGGACA TCTGGTCCTT TCCAATAATC TAGCTGAACG CGCCATTAAA TCATTGGTTA 1920

TGGGACGGAG TAAAAGAGTC CAGTGGACTC TTTTAGCCTA AGCTCAGTTT AAAAAAACGA 1980

GGGTGGTTAT TTTTAAAAAA GCGAGGGTGG TTATTTTCTC AAAGTTTTGA AGGAGCTAAA 2040

GCAAGAGCTA TTATTATGAG TTTGTTGGAA ACAGCTAAAC GTCATCAATT ATAGTGCGTT 2100

GAATCTATAA CAGTACGCAT CGACTGCTAA AATATTTCTA TAAATCAATT TTCCTTTCCT 2160

AATCGATTTG TTCATATCTT ATTACAATCC ATTATAAATA GCGAGAAATA TCTATCCTAT 2220

CTTCTAGAAT GTCTTCCAAA CGAGGAAACT CTCGTAAACA AAGAGGTTTT AGAGGCCTAT 2280

TTACCGTGGA CTAAAGTTGT ACAAGAAAAG TGCAAATAAG AAATCTCCAG ATTAGGAACT 2340

ATATATGAGT TCTCTAGTCT GGAGATTTTT CAATAGACTT CGTTATTGGG CGGTTACTTT 2400

CGAAACTTTG AAAACTTCAA AAAACGGATT TTTATCGCTC TGAACATCAA AAAAGAAAGG 2460

ACGAAATTTG TCCTTTCTCA AGCTTAGCTT TTCTTCAACC CACTACAGTT GACAAAGAGC 2520

CCTTTATTCT ATCAAACATG AAGCGCAAAA ACAAGCCAAA AATCCGATAG AATGGCTATC 2580

CCTCGACTAT CAAGTAAGAC ATTTCCATCA AATACGTTCA ATTTTACTCT TGTTCTACTA 2640

AGAATTAATC ATCTCGTTTT GATTTATTAA AAATATACAA TTCAGCTTTT CCTCCAAACT 2700

ATTTTATCCA CTATCCCTGT ATAGCTCTGT ATTATCTTAA CAACTTTAGT AGAGACATTT 2760

TCCTCAACAT AATCCGGAAC CGGTAATCCA AAATCCTCAT CTTGTGCCAA GCTAACAGCA 2820

GTTTCAACTG CTTGAAGAAG AGAATTTTCA TCAATGCCTG CCAAAATAAA TCCTGCCTTA 2880

TCTAAGGACT CAGGACGTTC TGTACTTGTA CGAATACATA CAGCGGGAAA AGGATAACCT 2940

TGACTAGTAA AGAAACTACT TTCTTCCGGT AAAGTTCCCG AATCAGATAC TACAACAAAT 3000

GCATTCATCT GTAAACAATT ATAGTCATGG AATCCTAGTG GCTCATGCTG AATCACACGT 3060

TTATCTAGTT TAAAACCGCT CTCTTGTAGC CTTTTCTTTG ATCTAGGATG GCAAGAATAT 3120

AAGATTGGCA TATTATACTT TTCAGCTAAT TGATTAATTG CTGTAAAGAG AGAAATAAAA 3180

TTTTTATCTG TATCAATATT TTCCTCACGG TGAGCTGAAA GTAAGATATA ACCTCCTTTT 3240

TTCAATCCCA AACGTTCATG GATATCTGAA GACTCAATAG CAGATAAATT TTTATGTAAC 3300

ACTTCTGCCA TAGGAGAACC AGTTACATAT GTGCGCTCTT TAGGTAAACC ACACTCATGT 3360

AAATACTTAC GTGCATGTTC AGAGTATGCT AAGTTAACAT CTGAAATAAC ATCAACAATC 3420

CGACGATTAG TCTCTTCCGG TAGGCACTCA TCTTTACAGC GATTGCCAGC CTCCATATGA 3480

AAAATTGGAA TATGTAAACG CTTGGCAGCA ATAGCTGATA AACAAGAATT TGTATCCCCT 3540

AAAATCAATA AAGCATCTGG TTTAATTTGA TTCATCAATT TGTATGAAGT ATTAATAATA 3600

TTCCCTACAG TAGCACCAAG ATCATCTCCA ACAGCATCCA TGTATACGTC CGGAGTGTCT 3660

AACCCTAAAT TATCAAAGAA AATACCATTT AAATTGTAAT CATAGTTTTG TCCAGTATGT 3720

GCCAAAATAA CATCAAAATA CTTTCGACAT TTAGTGATAA CACTACTTAG ACGTATAATC 3780

TCTGGACGTG TTCCCACAAT AATCAATAAC TTAAGTTTGC CATTATCTTT AAAGTGAATA 3840

TCACTATAAT CTGTCTTAAT TTTCATTTAT TTCTCCACTT GTTCAAAAAA AGTATCTGGA 3900

TGTCTAGGAT CAAATGACTC ATTAGCCCAC ATGACAGTAA TTAGATTTTC TGTATCAGAA 3960

AGATTAATAA TATTATGTGC ATAGCCCGGT ATCATATGTA TTGCTTCAAT CTTATCGCCC 4020

GACACTTCAA AGTTCAGAAT AGGATACTCT TGACCGTTTT CATCCAGCCC TATCCTACGC 4080

TCTTGTATTA AAGCACGACC AGAAACAACC ATGAAAAATT CCCACTTAGA ATGATGCCAA 4140

TGTTGCCCTT TGGTAATGCC AGGTTTAGAA ATATTAACAG AAAATTGACC CGTATTTTCT 4200

GTTTTTAATA ATTCCGTAAA ACTACCTCGT TCATCTATAT TCATTTTTAG AGGAAACTTA 4260

AACTTATCTA CTGGTAAATA AGATAGGTAG GTAGAATACA ATTTCTTTTT AAACGATCCC 4320

TGAGGAATTT CAGGCATAAC TAAACTATCA GGCTGTTTTT TAAATGTTTC TAATAGAGAG 4380

ACAATCTCTC CTAAGGTTGC ACGATGAGTC GTTGGTACGT AGCAGTAGTT TCCTGATGGG 4440

CTAGGTAAGA TTTGTAATCC ATCTAGATTA CAACGATGAG GATTTCCTTC CAATGCAGTT 4500

AGACACTCTT GTATCAAATC ATCAATATAC AGCAACTCCA ATTCTACACT TGGATCATTT 4560

ACTTGAATAG GTAAATCGTG AGCTAGATTA TAACAGAAAG TTGCTACAGC AGAATTGTAG 4620

TTAGGACGGC ACCACTTCCC ATAAAGATTC GGGAAACGGT AAACTAAGAC AGGTGCTCCC 4680

GTTTTCTTTC CATATTCAAA GAAGAGTTCT TCCCCTGCTA GCTTAGATTG TCCATATATA 4740

GAGTTTGAAA ATCGGCCTTC TAAACTAGCT TGAGTAGAAC TTGAGAGTAG AACAGGACAA 4800

GTGTTTTCAT ACTTTTCTAA AATCTCCAAT AATCTACTTG AAAAACCGTA ATTTCCCTCC 4860

ATGAATTCAT CAGGATTCTG TGGACGATTG ACACCAGCTA AATGGAATAC GAAATCGGCC 4920

TTCTTACAAT ATTCATCTAA TAAAATCGGA TCTGTATCAC GATCATACTG AAAAATCTCT 4980

CCAATCTCTA AATTAGGACG AGTCCTATCT CGTCCATCTT TCAAAGCTTC CAGAGTACAG 5040

ATAAGATTTT TTCCTACAAA TCCTTTCGCT CCTGTGATTA AAATATTTTT AATCATGCCC 5100

CCTCCTTATT TTATATGCTG TTTTAATAGT TAACTCTCTC GACAATACAT GATACATTAT 5160

ATATCCTTGA TAATTTTAAT GTATCTTAAA AGATTTTACA TCTCTTCGTC TGCTACCATA 5220

TCACGAATTG CTGTCTGTAT TTCATCTAAT TCTAGCAACT TTCTTTTAAC TTGCTCTACA 5280

TCCATCAAAT CGGTATTATT ACTATTGAAT TCTGTCAACA AATTTCTATT CGTACTACCA 5340

TCTTTGAAAT ACTTATCATA GTTAAGATTA CGATTATCAC TAGGAACTCT ATAAAAATCA 5400

CCCAAATCAA TTGCATTTGC GCACTCTTCG TTAGTTAATA GTGTTTCATA CCTTTTTTCT 5460

CCGTGTCTAA TACCTATAAT CTTAATATCT TGTTCTGAGG CAAAAATTTC TGATACAGCC 5520

TTAGCCAACA CTTCAATCGT ACATGCTGGT GCTTTCTGAA CTAGTATATC TCCAGATTTC 5580

CCTTCTTCAA ATGCAAATAA AACCAAGTCT ACTGCTTCTT CCAATGTCAT CACAAAACGT 5640

GTCATGCTAG GTTCAGTAAT TGTAAGAGCA TTTCCTTGCT TAATTTGCTC AATCCAAAGA 5700

GGAACGACAG ATCCACGGCT ACACAGAACA TTCCCATAGC GAGTCACACA TATCTTTGTA 5760

TGCTCAGGAT TTACCGTCCT GGACTTAGCA ACAGCAATCT TTTCCATCAT AGCCTTGGAT 5820

GTTCCCATAG CATTGACAGG ATAAGCCGCC TTATCTGTAG AAAGACAGAT AACTTGCTTT 5880

ACACCAGCTT CGATAGCCGC AGTGAGGACA TTCTCCGTTC CCAAAATGTT AGTTTTTACC 5940

GCTTCTACAG GGAAAAATTC ACAAGAAGGT ACTTGTTTAA GAGCAGCAGC GTGAAAAACA 6000

TAATCCACAC CATGCATAGC ATTTTTTACC GAAGCTAAGT CACGCACATC TCCAAGGTAA 6060

AAACGGATTT TCCCAGCCAC TTCTGGTACT TTTACCTGAA ACTCATGACG CATATCATCT 6120

TGTTTCTTTT CATCTCGCGA AAATATACGA ATCTCTGAGA CATCTGTTTC TAAAAAACGC 6180

TTGAGAACCG CATTCCCAAA TGAACCTGTC CCTCCTGTAA TTAGGAGAGT TTTTCCTGTA 6240

AATTGTGACA TATATTACAC TTCTCCTTCT AGTATGTCTG CAATTTTCTT ACAAGCCGTT 6300

CCATCTCCAT ATGGATTTGA AGCTTGACTC ATTGCTTGAT AAACTGAATC ATTTTCTAAT 6360

AATTCTTTAA AATGCCTATA AATATTATTT TCATCAGCAC CTACAAGTTT CAAAGTCCCT 6420

GCTTCAATTC CCTCTGGACG TTCAGTTGTA TCTCTCATAA CCAAAACAGG TTTTCCTAAA 6480

CTTGGAGCCT CTTCCTGAAT ACCACCACTA TCTGTTAAAA TTAAATAACT TCTTGATAAA 6540

AAATTGTGAA AATCTAATAC TTCTAAAGGT TCGATCATCT TGATACGTTC ACAGCCACTT 6600

AGTTCTTCCT CAGCAATTTG GCGAACACGA GGATTCATAT GGATAGGATA AATAGCCTTG 6660

ACATCTGAAT ATTCTTCAAT AATCCTTCTA ATTGCTCTAA ACATATGTCT CATCGGTTCA 6720

CCAAGATTTT CACGACGATG AGCTGTAATT AGAATAAACC TGCTTTCTCC TATCCATTCT 6780

AACTCAGGAT GCGTATAGTC CTCTTGAATT GTAGTTTGTA AAGCATCAAT CGCCGTATTA 6840

CCTGTCACAA ATATGCTCTC TGGAGTTTTT CGTTCTGTTA AAAGATTATC TTTTGAAAGT 6900

TGTGTTGGTG TAAAATGATA CTGAGCCAAA ACCCCAACTG CTTGACGATT AAACTCTTCA 6960

GGATATGGTG AATAGATATC GTAAGTGCGC AAACCAGCTT CAACATGACC AATTGGAATC 7020

TGTAAATAAA AGGCCGCCAG TGAACTAGCG AAGGTCGTAC TTGTATCCCC ATGAACTAAC 7080

ACCAAATCAG GTTTTTCTGA CTCTAAAATA GCCTTCATTC CTTCCAAAAT GCCAATGGTC 7140

ACATCAAATA AAGTTTGTTT ATCTTTCATA ATAGACAAAT CAAAATCGGG AATAATCCCA 7200

AATGTGTCCA AGACCTGATC CAACATTTGA CGGTGTTGGC CCGTAACGCA AACTAATGTT 7260

TCAATATTCT TACGTGTTCT TAACTCTTTG ACCAAAGGAC ACATCTTGAT GGCTTCTGGA 7320

CGAGTTCCAA ATACTACAAC TACTTTTTTC ATATATTTAC TTACTCCTAA CAAATAATGA 7380

ACGGTTCTTA AAATAAATTA GATAACGGCT AATCCATAAC ACCACCTCAG ACATACTTGA 7440

ACAAATAGCT AATGTTACTA AACTAAAATT ATCAGACAAG ATAAATATTC CTAATCCCAA 7500

AGTTTGGACA ATCGAAGCTA ATATAGTTGT CATTGTAGTT TCTTTCACTT TATCAATAGC 7560

TCCTAAGACA GGCCATCCGT AAATCATAGA ATAAAAACTA GCAACAAAAG CGGGTAATAA 7620

GTACTTAAGA AAATCTGCTG AAACGGTATA TTTTTCACCA CCAATTATAG AAAGAATTTG 7680

ATTTGAAAAG AATAAAACTA TCAAAACTCC AAAGATAATA GGAATAAACA TAATCCGATT 7740

AATACTCTTA ACCGATTGTA TATCTTTAGT ACGTATCATA TGCGGATATA AACTATTCGC 7800

TATAGGATTA TACAATGATT TTGCTGCTGA AAGCAGTTGC ATTGCTATCC CCCAAAAGGC 7860

TATCTCTTGA CTTTGTAAAT AAAAACCCGA AATGACTGTC GTAAAGACGC CAAAAATAGT 7920

AGTTGCAAAA TTGGATAAAA AATAAATAGA GGATTCCTTT AAATCTTTAA CCCAAACAGA 7980

CAGATAAGAA AATGATAATT TAATTCCATA ATAATGAAGG AATCTATAAG AAACTACTGC 8040

AGCAACTAAA TTCCCAATTC CTTCCAATAT AGGAATCCAT AAAATAGAAG AATCATCTTT 8100

TACTACAATA AATGTCAAAA TTGTAATGAT AGTTTTAGAA ATAATATAAG GAATTGCAAC 8160

TGCATGCATC TTTTCAATTC CACGAAATAA AAAGTCAAAG ATAAAAATAT TGGTCACTGT 8220

AGCTAACAAA TAAAAAACTG AAAAAAGAAT ATTCTCTCTC ATTATTGGGA TTTGCCACAT 8280

CAATATGGTG TAAATTAGAA TCGAAATGAT AGATAAAAAT ATTTTTTCAA CTAGAGTATC 8340

TCCAACTATC CTTCCAATCT TTGAGGGAGT AGTACAAGCA TTTACAATAT TTTTTGTAGC 8400

TGATATCATG AAACCAAAAT CAATCACCAG TTGAACATAA GCTATTAACG CTTTAACATA 8460

AATAACCATT CCATACGCGT CTAGCGAAAG CACCCTTGTC AAATACGGGA GTGTTAATAA 8520

AGGAAATAGT AATTTAACAA TATTCAGAAT ATAGAGAGAA CTTGTATTTT TTATAAATGA 8580

AATTCTATCA ACTTTCACGA ACTAGTCCTT CCAAAAAAAG ATCTAAATAG TCCAAACTAC 8640

TTCTCGCTTT CAACACCAAT TCTGAAGGTA TTGTTATCGG TTTTAGATGA AAAGTTTCAA 8700

GTTTCTTTAC AATACTATTA ACACTTGAAT CAAATAAAGA TTCACAACGT TGTAACTCTC 8760

CAATTGCTCC ATAATAACGT GCTGTTTTTT CTGGATGGCA TGCAATGGCA ATCACAGATT 8820

TATTAAAACA TGTTGCCACT ACCCCAACAT GTAATTTACA AGTTAAAACC ACATCTACCA 8880

TTTTCAACAA TGATGTCATT TCTGCAGGAG AATGATACTT GAATTGAAAA CAATCCTCAG 8940

TTCTAACTAA TTTTCTAAAT TCCTGATAAT AAGCATCTTC ATAAGGTAGA ATGGAATCCG 9000

AAGTTACTAC AACATAATAG TTAGGATTGT TTTCTAGAAA AAGACTAATT GATTCCGCAA 9060

ATTTTTCAAG AGCTTTTTTG GAATGATTAT AGTGAACAAG AATTATCTTC TTATCTTTAG 9120

CTTCTCTTTT CAATTGACAC AGCTGCTCTG TTTTTTCTTC TCTTAATTTA CTTGAAATAA 9180

TTAAATCAAA GGTTTCATGC ACTGGAGCCG AAGGCGACAA ATGCTTCAAA GAATCAAATG 9240

ATTCTCGATC ACGAACTGTA ATAAATTGAG CATGATTAAT AATTCTCTTT ATACCATAAT 9300

TCATCAAAGA ATCGTTATTA GGCCCTGCAC CAATACCTAA TACTCCTATA GGCTTTTTAA 9360

AATATGAAGC CCAAATTCCC AAAGGTAAAA ATCGTTTAAA TTGGATTAAA TTATCACGAA 9420

AACGTGCATT ATGCCCTTCC CCAAAATATC CTCCCGGGAT ATACAAAATA GCATCTGCTT 9480

GTTTTTTAGT AAAACTTTGT TTTTGGCGAT ATTCTTTCAA GTACATTTGA AAGAAATCTG 9540

ATGGATTATA AAAAGAAACT TCATATCCTT TAGATTCTAA TAAATCATAG ACAATCTCAC 9600

CGTAAAGATA ATCACCGTAA TTACTTGAAC CATAATCCGT TGCACCATGT AACATAATTT 9660

TTTTCACCAC TATTTTTTCA ACCTCCTAAA AATAAATATC ATAATCAAAC TATACATAAT 9720

AGGACGATAA ACATCTATTG AACTACTTCT CACTAAAAGC AATAGTTGAG AAATTACCGA 9780

AAAATAAATA ACTTTTGAGA TTTTACTTGT TTGAAAAGCT CTGAAATTTA ATCGCCATCC 9840

ACTAAATATT CCCAAAACAA AACTCCAAAA AACACCACCA TAGTAACCAA AGTTCCAAAA 9900

TAATTCTTCC ACAAAAGAAG AGCCTACAGG TAACCCCAAA AATTTATTAA TAACAACCGT 9960

CGCTGATGCT TTATCAAAAA AATCACCAAC TAACCATCCA ATAGGAAAAA TTGATAGGAT 10020

AGTCCGTAGA AATGTCATCC CATATTCATA TGGAATGCTA CTAGGCACAA CAGTTACAGC 10080

AGAAGCTACT GTTAGGCTGG TCAGTCCCGA CTCTGAAAAT ACTTCCCCTA GTATATTCTT 10140

TACAAAATCT AATGAAGAAA AGGAATCAAA TAAGTATATA CCTATAGTAT TCAAGTCGAA 10200

ACGGTGCCCC CTAATAACAA CTAATACATT TAATAGAAAT ACAGTTACTA TTAAAAATAC 10260

AAGTACTCTT TTCTTCGAAA AAGTAATCCC TAAAGATTGT GTGTATACTA AAACCAACGC 10320

CAAGATTGAA AACACCTGGA TTTTACGACT TCCTGTTAGG ATCATTATCA AAATTAGGTA 10380

AAACAACATT ACCCAAAAAA TAGTACGCTT TATAACTCGG GACAGCTTAT CTGAATAAAA 10440

CAAGGAGAAC ACACCAGGAA GCATAAGTAC TCCTAAATCA TCTATTATTC CTGAACTAGC 10500

TGCCTCTGAA TATGCTGAAT AGCTATTCGC CGCTCTAACT GCTAGTACTG TTTTAGAATC 10560

AGTTATTACC CTAGAAATAA AGCCCACTCC TGTTAAAATC CTACCCGCAT TGTACAAAAT 10620

TTTCTCTTCA TTTTCCTGAT AATTTTGTAC TTCTGAATGA TAATGTACCT TTCCATCACT 10680

ATAAAAAAAT AAATAGCCTA CAGAATAACA AAACAAAATC CAAATTATAA AAATATATGA 10740

ATGAAATAAT TCTTCATTAT TATAGAAGTT ACTAGGGCTC CACAGCAGAG TTGTTTGAAA 10800

CCCCATATAC TCATTGAAAA TTAATCCAAA CATAAAAAAA TAAGATAAAA TCAGATACCA 10860

TACAGAAAAA TCATATATAC TAACTTTTTG TAAAATAAAA CCAGTAATTT GAAAAATAAT 10920

TAGAAAGCAA ACCCATATAA ATATAGACGG AACATAATTA GATATAAGAA AACCATTATT 10980

CCAATTATCG AGAGTCCAGA ACAAGTAACA GAAAGCAAAT ATAAAACTTA ATGTCACTAG 11040

TGTCACTCTA CAAATATACT TTGTCTGCAT CTATATCTCC TTTATTACAC ACATTTCTTG 11100

ATAACGATTC AATAATTTAC TAGCTTGATA ACAAATATCA TAGAGTCCAT CTGTCATACT 11160

GTTATTTATT TCAAAACGAT TGCATTCCTC AGATGTTAAA GACAGTACTT TATCTTTCCA 11220

TAGCAACACA GACTCTTCGT TGATAGGTAA GTAACTAATG TTTTTGGTCA CATCTACTTC 11280

TTGCGTCACT GTATCTGACG ATAAAATTTG TAATCCCGAT GCCTGAGCCT CTACTAGAGA 11340

AACAGGCAAC CCCTCATATT TAGACGGAAG CAAAAAAACA TCCATCGCAG ATAATAAATC 11400

AGAAATATCA GTCCTTCTCC CTAAAAATAG CACATATGGG GTCAGATTTA GTTCTAAAGC 11460

TTTCTGTTTT AATTTCTGCT CATCCTCACC ATTACCAACT AGGAGTAAAA TAACATTTGG 11520

TTTGATTAAA ATGAGTTCTT TTAAAACGTT AAATAAATAA CTTTGGTTTT TTTGATCTGA 11580

TAGGCGAGCT ATATTTCCTA ATACGAACTT ATTTGACACA TCTAATTCTC TACGACATTT 11640

TTCTCTAACA TCTGACAAAA ATTGATACTT TTTCAAATCA ATTGCATTAA AAATAATTTC 11700

AATTTTTCCG TCTTTATACG CTTTCTCTCC ATATAACCAC TTAGCCGAAT CTTCCCCACA 11760

TGCAAACCAA TGAGTTGCTA AGATTTTTAC CAAAATTGTT ACTAATTTAC GCAATACTTT 11820

TTGAAAACTG TTTTCTGTTA CATAAGCCAT ATGACTATGA ATAATTCTAA TTTTACAACC 11880

AATTATTTTA GATAAGATCA GACCAATTGC AGATTTATAG CCATGGCAAT GAACTATATC 11940

ATAATCTCCT TTCTTTATTA TTCTAGCAAG AGAGAGAAAC TGATGTAGAG GCTTTTTCCT 12000

TAATAGAGGC ACATGATAAA CCTTTGCACC CAATTCTTTC ATTTTATCCT CTAAAAATCC 12060

TTGTTCTTTT CCAGGCACAA TAAAATCAAA TTGAATTTTT TTTCTATCAA TGTGAGAATA 12120

ATAGTTGAAT AGAAAACTTT CTACTCCACC ACTATCTAGT GTTGTAAATA GATGTAATAC 12180

TTTAATCATT CTTCTTCCTT AAGCTTAAGA TTCGCTTCTC TAATTCTATT TCTGTTTTTT 12240

GTTTTTCTAA ACTAATTCTG TCCATGAAGT TATCACAATT CTTAATTAGC TGTTTCCTGT 12300

CAAGGTTTTG AATATACAAA GCCAAACAAT CTTTTTCCGA TTCATCCTTC ATAGGTAAAA 12360

CGAAACCAAA ACCATTCTCT ATTGACACTT TTTCCATATA AGTATCTTCA CAAACTAAAA 12420

TAGGTTTATA CAACAATGCA GCAAAGTAGA GTTTATTAGA CAAAGCATAG TCTAGTAAGG 12480

GAGTGTGATT CCCGTATAAA TTCAAAACAA CATCTGTATT CTTATAAAAA GACATGGTAT 12540

CTTTAGGCTG GAATGTGTCC ACCAAGTTAA CATTGCTGAT ATTTTTTTCT TGACAAAATT 12600

CCCTTAATTC TCCTGCATTA GTACCTATAA AATTCAACTG AAATCGACTG TCATTTGCAA 12660

AAAAATCGAT TATTTTTTTA TTTTGTTCTT GAAAACGAAT TAAACCAATG TAGGAAAGTT 12720

GAATTGGAAA CGTACTATTA TTTTTTAACT GCTTTACCTC GTTTAATTCT ATCATATTGG 12780

GTAGGTTATG GGTAGTAAAA TACTCTCCCA TTGGTAAAAA AAATTTATAG CCGTCTGAAG 12840

AAACGATATT CATTAAAGAA TTTTTCACCA ATTGTTTCTG AACCAAACGA TAAACCAAAA 12900

ATTTTTCATA ACTGTAATCA CGAATATCAT AAATATATCT ATTTTTAAAT GAAAAGAGAA 12960

GAAAATCTAC TAAAATGAAA GACACAATAC TATGTAACGG CAATATCATA TCATAATCAT 13020

TTTCTTTTAG CTTCTTTTTA ATTTCTTTTC TGAATTTTAC ATAACCTAAT ATCTTACTTA 13080

ATTTTCCTTT ACCAGAAAAA GAAATACGAT AGTAGTTTTG TTTTGTAATA ATCTCGTTAA 13140

TATTCTTATC CCAATATATA ACATCGTAAC TAATAGACAG TTTCTTCAAT AATTCTTTAT 13200

AAAAATTGAA GTAAGGAGTT AGATATATAT TATCAGATAG TATAAACAGT ACTCTCATTA 13260

AATTATTCTT TCTTACTTTC CCTCTCTAAA CATGTCTCCA GTTCGAGCAT AAACTGCTCT 13320

TTTGAAAAGT GATTTTCATA GTAACAACGA GCTTTCTTTC CTAACTCTCT TTGTCTCTTA 13380

ATAGATAACA TACTAAATTT ACAAATATTT TTTGCCAATT GTTTTACATC TCGTTCGGGA 13440

CTAACATATC CACAATTTGC TTCTTCTACA ATTATTTTAG CATCTCCTGA AATTGCACCT 13500

ATAATTGGTT TGCCTGCCGC CATATAAGAk TGTACCTTCC CAGGTATAGT ACGAGAAACT 13560

ATCGAGTCTC CTATTAAAGA AACTAACATA GCATCTGATT TTTTATAGAA GGATGGCATT 13620

TCCTCCAAAG AACGTCTTCC ATAGAAGGAA ATATTCTTTA ACTCCAATTC ATGAGCTAAT 13680

GCTTTCATGC TTAACAATTC CGTACCATCT CCAACAAAAT GAAAATGAAT TTTCTTGGGT 13740

AAATTGGTAT TCTTCTCTAT CAAACTGGCA GCTTTCAAAA TAGTTTCCAA ATTTTGTGCT 13800

TTGCCAATAT TACCAGCAAA AGTTAGGTCA ACACTTTCTT TATTAACTAT AGATTCATCA 13860

GGGATAAAAA GATCTTCTGC ATATTGTGGC AAATATGTAA TCTTTTGTTC GGATATGTCA 13920

AATTGCTTCA CAAAATAATT TTTAAATGAT GGACTAGTGA CAAATATATA ATCACTAGCT 13980

CGGTAAACTT TTTTTGAGAT AAATTTAAAC AGCTTGAAAA TCAAGCCATC TTGTTTCACT 14040

CCACCTACGG TTAAACTATC TGGCCAAACA TCCATACAAT ATAGAAACAT CGGTTTCTTA 14100

TATTTTTTTT TATAAGCCAT ACCAGCCCAT GCCATCATAA CTGGAGACAA TTGGTTAACG 14160

AATACACAGT CAAAATTCGA TCCATCTTTC GTTTTATACC TCCCCAATAA AACTCCTAAA 14220

GTAGAACTAA TTGCAAAGCT AAAATAATTC AACAATCGAA ATACAACACT TTTTTTTCTA 14280

GGGATTGTAT AAGAACGATA TATCGTAACA CCTTCTATAA TCTCACGTCT TTTTTTATTA 14340

TGACGATAAT CTGCATATAT CTTCCCTTCA GGGTAATTAG GAATCCCAGC CAAAACAGAG 14400

ACTTCATGCC CTTTTCGAAC TAAATCTTCA CAAATATCTG ACAACCTGAA TGGTTCTGGC 14460

TTATAATGTT GGCAAACAAA TAGTATTTTC ATTGTCCAAT TTAACTTTCT TTCTTACCAC 14520

TACCCTCTAC AATACCTTTT CGTTTCAGTA CGTAAGGTAT TGTCTTAACT ATACATCTAA 14580

TATCCATTAT CAAAGACAGA TGTTTAACAT AGTAGCCATC TAACTCCGTC TTCATCTCAA 14640

CAGACAAAGT ATCACGCCCG TTAATTTGTG CCCATCCAGT TAACCCTGGC AAGATATCAT 14700

TTGCTCCATA CTTATCTCTC TCTGCAATCA AATCTAGTTC ATTTATACCC GCTGGTCTAG 14760

GACCTACAAT ACTCATATTA CCAACAAGAA TATTAAACAA TTGTGGTAGT TCATCCAAAG 14820

ATGTTTTTCG CAAGAAAGCC CCTACTTTTG TAATCyATTG CTCTGGATTA TATAAGTTTC 14880

GAGGCGCCAC ATTTTTAGGT GCATCTATTT TCATAGACCT AAATTTCAAA ATATAGAAGT 14940

ATTCTTTATG AATACCAAAG CGTTTTTGCT TAAATATAAC CGGACCTTCT GAATCAAGTT 15000

TAATCGCAAT TGCAATTATC ATAAAAACCG GACACAATAT TATTATCCCT ATTAAAGATA 15060

ATAATATATC ACCTAATCGT TTTATTATAC CGTACATAAA CAACCTCCAA CTATAAATTC 15120

TATTTCCATT TTTCATTCTA TTTCCATTTG ACAAATTAAA TCAGGCAGTA CATGCAACTA 15180

CAGAAACTCA ATATATATTT GGTCACTCAA TGATTTTCAG AAATATAATT CTTTTATCCT 15240

CTACGTCAGA TAAAACTTTT CTCCATCTAA ACAAAATTTA TTTGTTTCAG TAATATATGA 15300

GTTCTCAATA ATGAATTAGA AGGTCCAGTT CAATTATTCT TCCAAATAGA CCGAATATTA 15360

TTTGAAGACA TATCGGTTTC TGAAATTGCA ATCAGTACAT AAGCTAATAA ACTGATAAGT 15420

ATGCTCTGTA AGAATGCCAG AGTTATATTG TAGTCCCCTT CCATACTATA TTCATTTTAT 15480

TTTTTACCAT AATTTCCATA GGAACCGTAA ACTCCATACT TATTAACCGA GATATCCAAT 15540

TTATTTAAAA CAACTCCTAG GAACAGTTTC CCTGTTTGTT TTAATTGTTG TTTCGCTTTT 15600

TGGATATCAC GTTTATTCGC CTCACCTGTT GCTGTTACCA AGATGGACGC ATCACACTTT 15660

TGAGTGATAA TTGCCGCATC AATAACAATT CCAATAGGCG GTGTATCAAT AATGATATAA 15720

TCAAAATATT TACGCAATGT TTCAATCATA TCATTAAAAT TTTTACTTTG TAACAAGGCT 15780

GTAGGGTTTG GTGATACAGA TCCCGATTGA ACTACAAATA AATTTTCAAT ATTTGTATCA 15840

CATAAACCGT GAGATAAATC AGCTGTCCCA GATAAAAATT CTGTTAGCCC TGTAATTTTT 15900

TCACGAGATT TAAAAACTCC TAACATAACT GAATTTCGAG TATCGCCATC GATCAAAAGA 15960

GTTTTATAGC CTGCACGCGC AAACGACCAT GCTATATTTA TGGAAGTAGT TGTTTTTCCT 16020

TCCCCAGGGT TAACAGAAGT AACGGAAATT ACTTTTAGTT TATCTCCGCT CAACTGTATA 16080

TTTGTACACA AGGCATTGTA ATATTCTTCT GCCTTCTTAA TGAACTCCAG TTTTTTTTGT 16140

GCTATTTCTA ATGTCGGCAT CCTTCTCTCC TATTTCAACT TACCCAAGTT TGGCACAACT 16200

CCCAAAAGTG TCATCTGCAA TGTATTTTCG ATATCTTCCG GACGTTTCAC ACGAGTATCC 16260

AAAAGTTCAA GATGAAGAAC TATAACACTA GTTCCAATCA CCCCTGCCAA AAAACCAATT 16320

AGTGTATTGC GTTTAATATT TGGCGAAGAC GGGGATATCG CCGGCCTTGC CTCCTCCAGT 16380

GTTGTCACGT CAGAAACACG AGTAATACTG ATAATTTTTT GAGCAGCTAC TTCTCTCAAA 16440

GAGTTAGCGA TACGGCTTGC CTCTTCAGGA ACTCGATCAT TAACTGAAAT AGAGACAATA 16500

CGGGTATCAA CTGGTACTGT CACTTTAATT TTATTAGCCA AACCTTTTGG CGTCAAATCT 16560

AGTTTCAAAT CAGAAACAAC TTCCTCCAAA ACATCCTGCG AAAGGATAAT CTCACGGTAG 16620

TCTTTTACCA GATAAGTTCC TGCCTGCAAA TCCTGATTTG TCAACCCCGG CTTGTCTCCT 16680

TGATTGCGAT TCACTACGTA AATTCGCGTG GTACTCGTAT ATTCTGGCTT AACAATAAAA 16740

GTGCTATATG CAAAAGCCCC CGCACCTGTC ACAAGTGCCA CTATTAAAAT CATTAGCTTG 16800

CGTTTCCACA AGCTTTTAAC TAATTGAAAT ACATCGATTT CTATCGTATT TTGTTCTTTC 16860

ATCATTTCTC CTAAATTAGT TGATCCATTA CAATTTTTCG AGGATTGTCT ATAAAAAGTT 16920

CCTGAGCCTT CGCTTCTCCG TATTTTTGGG TAACAAGGTC ATATGCTTCT GCCATATGAG 16980

GAGGTCTACC GTCTAGATTG TGCATATCAC TTGCAATGAC ATGAACCAAA TCCTGCTCTA 17040

AAAAATACTG AGCTCTTTTT TTCATGAATT TATAACGTTC GCCAAAAAGT TTGGGTTTGA 17100

GGACATGTGA ACTATTTACT TGCGTGTAAC AGCCCATATC GATCAGTTCT CGAACGCGTT 17160

TTTCATTATT TTCAAGAGCA TCATAGCGCT CAATGTGGGC AATGACTGGA GTAATTCCCA 17220

ACATCAAGAT CTTGCTCAAG GCGCTATGAA TATCGCGATA AGGAGTGTTC ATACTAAACT 17280

CTATCAAGGC ATAACGACTA TCATTGAGGG TCGGAATCCG CTTTTTTTCC AGCTTATCCA 17340

GAACATCTGG TGTGTAATAA ATTTCAGCCC CGTAAGCAAT GACCAAGTCA CTCGCCACTT 17400

CCTTAGCTAT TTCCCGAACC TGAAGAAAGT TTTCTGCTAT CTTCTCTTCC GGAGTTTCAA 17460

ACATGCCCTT GCGACGGTGA GAGGTAGAAA CAATGGTTCG CACCCCCTGT CTGTAGGATT 17520

CTGCCAAGAG AGCCTTGCTT TCCTCTCTTG ACTTGGGACC GTCATCTACA TCAAAAACGA 17580

TATGCGAATG GATGTCTATC ATTTCATCTA CCCTCCATCA CATCCTGTAT AGCTGCTTTA 17640

ACTACAGCTA AACTACTATC ATCTATTTCC ATCACATAGA GGTTACTGTC TGGCATTGCA 17700

TAAGAAGGAA GATCCATCCG ACCTGTCCCT TTTAAATCTT GAGAATTTAC TTTATAATTC 17760

CCTCCACTTT CTAACTGAGC ATTGACCAAA TTTATCATGG TCTCAAGTGG CATATTTGTT 17820

TGGATAGAAT CTTGCAAGCT ATTAATGATC GTACTATAAT TTTTCAGCAC TTCGGTTGAC 17880

GTTAATTTTT GAAGGATAGC CACAATCACC TTTTGTTGAT GGCGCCCGCG GTCACGATCG 17940

CCATCTGCTA GGGAGTAGCG CTCACGAACA AAACCGAGAG CCTGTTCTGA ATCAAGATGA 18000

ACATTGCCTG CAGGGTAATA CTTTCCATTC GTATGGGCAG TAAATTCTTG ATCATTATAA 18060

ACATCAATTC CACCCAACAA ATCAATCAAT TTCAAAAACG AAGTGAAGTT CAATCGCACA 18120

TAGTAATTGA TATCCACTCC ATAGAGATTT TCTAAGGTGT GAATGGACGA ATCAACTCCA 18180

TAAATGCCCG CATGAGTCAA TTTATCTTTT TGATTATTTC CACCATCTGC GATTGGTACA 18240

TAGGCATCAC GTGGCGTTGT GGTCAAGAGG ATTTTCTTGG TATCTCGATT GACAGTCATC 18300

AGGATGTTGA CATCTGATCG CGACACCGAA CTAATAGGAC CATAGGTGTC AATTCCACTA 18360

ACATAGATAT TGAAAGACTG ACTCTTAGAC GTCTTAGGAG CTTCTACTTT TTTAGTGAAT 18420

CCCTTAGTAT AAATCTTTTT TATCTTCGAT GCGTAGTCTG GATACTCTGA CTCGATGATG 18480

TTTTCAAAGA CACTATTTAG GACAATGGCC TTAGTCTCCC CTGCAATCAA ACTCTTGTAA 18540

GCTGCCAAGT AAGACGAACT CTGGTTGACC GTCAAATCGG TATTCTGACT TGACTTGATA 18600

TCAGCTAGTA ATTTCTGAAT ATTTTCATTA TTAGTCCCAG TCGGTGCTGT CACACTCGTC 18660

AGTTGCGTAA CATTTTCGAT CTCACTATCT GCTAAAACAG CGACACTGAT TGAATATTCT 18720

GAGTAATTAG AAGTCGCATT TAAACGATTG GTCAGTCCAA CAAACTGCTG TACTGCAAAG 18780

AGCGACACAG AGCTGACAAG GATAGAGAAC ACCAACAGAA AAATAGTAAA CTTTTCAGCT 18840

TTTTTATAGA TAATCAAGAG TAGCCCTACC AAGGCAACTA GTAGGACTAA CGCAGTTACC 18900

ACTAGATTAA GATATCTAAA AGCAAGGATA TTGTACTTAA AGATTAAGAA CAATAAAAAA 18960

CAAACTAACA ATAAATAAAT AGTCAGCAAA ACTATATTAA CACTTCGCTT CACTTTCTGT 19020

GAACGTGATT TTTTAAAACG TCTACTCATG ATTAATACCT ATACATTGAA CATTATACGA 19080

TTATATCACT TTTTTACGGT AATGTCTACA CCTTTATTTT TACTATCTGC ATCTTTAAGT 19140

ATCTTAGTAG ACTTCCCGCG AAACAAAAAT ATAGTAAAAT GAAATAAGAA CAGAACAAAT 19200

CGTTCAGGAC AGTCAAATCG ATTTCTAACA ATGTTTTAGA AGCAGAGGTG 19250

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21706 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

AAAGTTGAAA GACTGCTAGC TGTTTTTGAT ACCAATCGTT TCCAACTACA GAGCAAACAG 60

TATACAAAGT TTGTTTTTGG ATGTAAGCTT CTTGATGGAC AATTCCAAGA AAATCAAGAA 120

ATTGCTGACC TTCAATTTTT TGCCATTGAC CAACTGCCGA ACTTATCTGA AAAACGCATT 180

ACCAAGGAGC AAATAGAGCT TCTTTGGCAG GTTTATCAAG GTCATAGGGG GCAATATCTT 240

GACTAAGAAG ATGATTATCG TATTTCTAAA TCCATTTTTA ACAACTAGCA TGGTATAATA 300

ATATGCAGGA AAATTTTGAA TTATGAGGAA GACTAGATGA ATTTATGGGA TATTTTCTTT 360

ACGACTCAGG CAACCGAGCC GCCCAAATTT GACCTTTTTT GGTATGTTAG CCTATTTACG 420

CTCTTAGCCT TAACCTTTTA TACAGCCCAT CGCTATCGTG AAAAGAAGGT TTACCAACGA 480

TTTTTCCAAA TCTTGCAGAC TGTTCAGTTA ATCCTTCTTT ATGGTTGGTA CTGGGTCAAT 540

CATATGCCAC TGTCAGAAAG CCTACCCTTT TACCATTGCC GTATGGCTAT GTTTGTGGTA 600

CTCTTGCTTC CTGGTCAATC CAAATATAAA CAATACTTTG CATTATTGGG AACATTTGGG 660

ACATTAGCAG CCTTTGTTTA TCCAGTGCCA GATGCTTACC CTTTTCCACA TATCACCATT 720

CTATCCTTTA TCTTTGGTCA TTTAGCACTC TTGGGGAACT CTCTAGTTTA TCTATTGAGA 780

CAGTATAATG CGCGATTGCT GGATGTGAAG GGAATTTTTC TCATGACCTT TGCCCTAAAT 840

GCCTTGATTT TTGTGGTCAA TTTGGTGACA GGTGGCGATT ACGGATTTTT GACAAAACCG 900

CCATTGGTTG GGGATCACGG TCTAGTAGCT AATTATTTAC TTGTTTCAAT TGTGCTGGTA 960

GCTACTATCA GTTTGACTAA GAAAATCTTA GAATTCTTTT TAGCTCAAGA AGCAGAAAAA 1020

ATGATTGCAA AGGAAGCTTA ACACAGAGCT TTCTTTTTTG CTCTTAGAGA GTTTTTACAA 1080

GCAGCTTATA AAATAAGAAT TTCTGAATAG ACAAACTCAA AAAATGGCTG GGAAATTTAG 1140

GAAAAAAGCA AGCACGATTA AATTTTTTGT GTTATAATAT TTTGTGAATA GCTATGCCTA 1200

TGTTTAGCTA TGGAATAATA CGAAGTGCGA AACTTGGAAG ATAGAGAGGA AGCGATGTAA 1260

TGGCTAGAGA AGGCTTTTTT ACAGGTCTAG ATATTGGAAC AAGCTCTGTC AAGGTGCTTG 1320

TGGCCGAGCA GAGAAATGGT GAATTAAATG TAATTGGCGT GAGTAATGCC AAAAGTAAAG 1380

GTGTAAAGGA TGGAATTATT GTTGATATTG ATGCAGCAGC AACTGCTATC AAGTCAGCCA 1440

TTTCCCAAGC GGAAGAAAAG GCAGGCATTT CGATTAAATC AGTGAATGTC GGCTTGCCTG 1500

GTAATCTTTT GCAGGTAGAA CCAACTCAGG GGATGATTCC AGTAACATCT GATACTAAGG 1560

AAATTACGGA TCAAGATGTT GAAAATGTTG TCAAATCAGC TTTGACAAAG AGTATGACAC 1620

CTGACCGTGA AGTCATTACC TTTATTCCTG AAGAATTTAT TGTGGATGGT TTCCAAGGGA 1680

TTCGTGACCC ACGTGGCATG ATGGGGGTTC GCCTTGAAAT GCGTGGTTTG CTTTATACAG 1740

GACCTCGTAC TATCTTGCAC AATTTGCGTA AGACGGTTGA GCGTGCAGGT GTTCAGGTTG 1800

AAAATGTTAT CATTTCACCA CTAGCAATGG TTCAGTCTGT TTTGAACGAA GGGGAACGTG 1860

AATTTGGTGC TACAGTGATT GATATGGGGG CAGGTCAAAC GACTGTCGCT ACAATCCGTA 1920

ATCAAGAACT CCAGTTCACA CATATTCTCC AAGAAGGTGG AGATTATGTA ACTAAAGATA 1980

TCTCCAAGGT TTTGAAAACC TCTCGCAAAT TAGCGGAAGG CTTGAAACTG AATTACGGGG 2040

AAGCCTATCC GCCTCTTGCA AGCAAAGAAA CCTTCCAAGT AGAGGTTATT GGAGAAGTAG 2100

AAGCAGTCGA AGTGACGGAA GCCTACTTGT CAGAAATTAT TTCTGCACGA ATCAAGCACA 2160

TCCTTGAACA AATCAAGCAA GAATTAGATA GAAGGCGTCT ATTGGACCTC CCTGGTGGTA 2220

TTGTCTTAAT CGGTGGGAAT GCCATTTTAC CAGGTATGGT TGAGCTTGCT CAGGAAGTCT 2280

TTGGCGTCCG TGTCAAGCTT TATGTTCCAA ATCAAGTTGG TATCCGTAAT CCAGCCTTTG 2340

CGCATGTGAT TAGTTTATCA GAATTTGCGG GTCAATTAAC AGAAGTTAAT CTTTTGGCTC 2400

AGGGAGCGAT AAAAGGTGAG AATGACTTAA GTCATCAGCC AATTAGTTTT GGTGGGATGC 2460

TGCAAAAAAC AGCTCAGTTT GTACAATCAA CGCCTGTTCA ACCAGCTCCT GCTCCAGAAG 2520

TAGAGCCGGT GGCGCCTACA GAACCAATGG CGGATTTCCA ACAAGCTTCA CAAAATAAAC 2580

CGAAATTAGC AGATCGTTTC CGTGGATTGA TCGGAAGCAT GTTTGACGAA TAAAGAGGAA 2640

AAATAAATTA TGACATTTTC ATTTGATACA GCTGCTGCTC AAGGGGCAGT GATTAAAGTA 2700

ATTGGTGTCG GTGGAGGTGG TGGCAATGCC ATCAACCGTA TGGTCGACGA AGGTGTTACA 2760

GGCGTAGAAT TTATCGCAGC AAACACAGAT GTACAAGCAT TGAGTAGTAC AAAAGCTGAG 2820

ACTGTTATTC AGTTGGGACC TAAATTGACT CGTGGTTTGG GTGCAGGAGG TCAACCTGAG 2880

GTTGGTCGTA AAGCCGCTGA AGAAAGCGAA GAAACACTGA CGGAAGCTAT TAGTGGTGCC 2940

GATATGGTCT TCATCACTGC TGGTATGGGA GGAGGCTCTG GAACTGGAGC TGCTCCTGTT 3000

ATTGCTCGTA TCGCCAAAGA TTTAGGTGCG CTTACAGTTG GTGTTGTAAC ACGTCCCTTT 3060

GGTTTTGAAG GAAGTAAGCG TGGACAATTT GCTGTAGAAG GAATCAATCA ACTTCGTGAG 3120

CATGTAGACA CTCTATTGAT TATCTCAAAC AACAATTTGC TTGAAATTGT TGATAAGAAA 3180

ACACCGCTTT TGGAGGCTCT TAGCGAAGCG GATAACGTTC TTCGTCAAGG TGTTCAAGGG 3240

ATTACCGATT TGATTACCAA TCCAGGATTG ATTAACCTTG ACTTTGCCGA TGTGAAAACG 3300

GTAATGGCAA ACAAAGGGAA TGCTCTTATG GGTATTGGTA TCGGTAGTGG AGAAGAACGT 3360

GTGGTAGAAG CGGCACGTAA GGCAATCTAT TCACCACTTC TTGAAACAAC TATTGACGGT 3420

GCTGAGGATG TTATCGTCAA CGTTACTGGT GGTCTTGACT TAACCTTGAT TGAGGCAGAA 3480

GAGGCTTCAC AAATTGTGAA CCAGGCAGCA GGTCAAGGAG TGAACATCTG GCTCGGTACT 3540

TCAATTGATG AAAGTATGCG TGATGAAATT CGTGTAACAG TTGTTGCAAC GGGTGTTCGT 3600

CAAGACCGCG TAGAAAAGGT TGTGGCTCCA CAAGCTAGAT CTGCTACTAA CTACCGTGAG 3660

ACAGTGAAAC CAGCTCATTC ACATGGCTTT GATCGTCATT TTGATATGGC AGAAACAGTT 3720

GAATTGCCAA AACAAAATCC ACGTCGTTTG GAACCAACTC AGGCATCTGC TTTTGGTGAT 3780

TGGGATCTTC GCCGTGAATC GATTGTTCGT ACAACAGATT CAGTCGTTTC TCCAGTCGAG 3840

CGCTTTGAAG CCCCAATTTC ACAAGATGAA GATGAATTGG ATACACCTCC ATTTTTCAAA 3900

AATCGTTAAG TAAATGAATG TAAAAGAAAA TACAGAACTT GTTTTTCGAG AAGTTGCAGA 3960

GGCTAGTCTG AGTGCTCATC GAGAGAGTGG TTCGGTCTCT GTCATTGCAG TTACCAAGTA 4020

TGTAGATGTA CCGACAGCGG AAGCCTTGCT TCCGCTAGGT GTCCATCATA TCGGTGAAAA 4080

TCGTGTAGAT AAGTTTCTGG AAAAATATGA AGCTTTAAAA GATCGAGATG TGACTTGGCA 4140

TTTGATTGGT ACCTTGCAAA GACGTAAGGT GAAAGATGTC ATTCAATACG TTGATTATTT 4200

CCATGCATTG GACTCAGTAA AGCTAGCAGG GGAAATTCAA AAAAGAAGTG ACCGAGTCAT 4260

CAAGTGTTTC CTTCAAGTAA ATATTTCTAA AGAAGAAAGC AAACACGGTT TTTCGAGAGA 4320

GGAACTGCTG GAAATCTTGC CAGAGTTAGC CAGACTAGAT AAGATTGAAT ATGTTGGTTT 4380

AATGACGATG GCACCTTTTG AGGCTAGCAG TGAGCAGTTG AAAGAGATTT TCAAGGCGGC 4440

CCAAGATTTA CAAAGAGAAA TTCAAGAGAA ACAAATTCCA AATATGCCTA TGACCGAGTT 4500

AAGTATGGGA ATGAGTCGTG ATTATAAAGA AGCGATTCAA TTCGGTTCCA CTTTTGTTCG 4560

TATAGGTACA TCATTTTTTA AGTAGGAGAG AACCATGTCT TTAAAAGATA GATTCGATAG 4620

ATTTATAGAT TATTTTACGG AGGATGAGGA TTCAAGTCTC CCTTATGAAA AAAGAGATGA 4680

GCCTGTGTTT ACTTCAGTAA ATTCTTCACA GGAACCGGCT CTCCCAATGA ATCAACCTTC 4740

ACAGTCGGCT GGCACAAAAG AGAACAATAT CACCAGACTT CATGCAAGAC AACAGGAATT 4800

GGCAAATCAG AGTCAGCGTG CAACGGATAA GGTCATTATA GATGTTCGTT ATCCTAGAAA 4860

ATATGAGGAT GCAACAGAAA TTGTTGATTT ATTGGCAGGA AACGAAAGTA TCTTGATTGA 4920

TTTTCAGTAT ATGACAGAGG TGCAGGCTCG TCGTTGTTTG GACTATTTGG ATGGAGCTTG 4980

TCATGTTTTA GCTGGAAATT TGAAAAAGGT AGCTTCTACC ATGTATTTGT TGACACCAGT 5040

GAACGTTATT GTAAATGTTG AAGATATCCG TTTACCAGAT GAAGATCAAC AGGGTGAGTT 5100

CGGTTTTGAT ATGAAGCGAA ATAGAGTACG ATAATGATTT TTTTAATTCG TATGATTTAT 5160

AATGCAGTGG ATATTTACTC CCTGATTTTG GTAGCCTTCG CTGTCATGTC TTGGTTTCCA 5220

GGTGCCTACG AATCCAGTTT AGGTCGTTGG ATTGTAGCGT TGGTGAAACC AGTGCTTGCT 5280

CCCTTGCAAC GCCTGCCTTT ACAGATAGCG GGTCTTGATT TATCTGTTTG GGTTGCGATT 5340

GTTTTGGTTC GATTTTTAGG AGAAAACCTA GTGCGTTTTC TGGCGATGAT AGGATGAATA 5400

AAGGGATTTA TCAGCATTTC TCCATAGAAG ATCGTCCATT TCTTGACAAG GGAATGGAAT 5460

GGATAAAGAA GGTAGAAGAT AGCTATGCTC CTTTTTTAAC TCCTTTTATC AATCCTCATC 5520

AGGAGAAGCT ATTAAAGATT TTGGCCAAAA CCTATGGTCT TGCTTGTAGC AGTAGTGGGG 5580

AATTCGTCTC GAGTGAGTAT GTTCGAGTTT TATTATACCC AGATTATTTC CAACCAGAGT 5640

TTTCAGATTT TGAAATATCT CTCCAGGAAA TTGTGTATTC CAATAAATTT GAACATTTAA 5700

CGCATGCTAA GATTTTAGGG ACAGTCATCA ATCAATTAGG GATTGAACGG AAACTTTTTG 5760

GAGATATCCT AGTAGATGAA GAACGGGCGC AGATTATGAT TAATCAGCAG TTTCTTCTTC 5820

TCTTTCAAGA TGGACTAAAG AAAATTGGTC GTATACCTGT TTCGCTGGAG GAACGTCCTT 5880

TCACCGAGAA AATAGATAAG CTAGAACAGT ATCGAGAACT GGATTTATCT GTGTCTAGTT 5940

TTCGATTAGA TGTTCTTTTA TCAAATGTTT TGAAACTATC TAGGAATCAA GCAAACCAGT 6000

TGATTGAAAA GAAACTTGTC CAAGTAAATT ATCATGTGGT AGACAAATCA GATTACACTG 6060

TTCAAGTTGG AGACTTGATT AGTGTGAGAA AATTTGGTCG CTTGAGATTA CTTCAAGATA 6120

AGGGACAAAC GAAAAAAGAG AAGAAAAAAA TAACCGTCCA GTTATTATTA AGTAAGTGAG 6180

GAATAGAATG CCAATTACAT CATTAGAAAT AAAGGACAAG ACTTTTGGAA CTCGATTCAG 6240

AGGTTTTGAT CCAGAAGAAG TCGATGAATT TTTAGATATT GTGGTTCGTG ATTACGAAGA 6300

TCTTGTGCGT GCGAATCATG ATAAAAATTT GCGTATTAAG AGTTTAGAAG AGCGTTTGTC 6360

TTACTTTGAT GAAATAAAAG ATTCATTGAG CCAGTCTGTA TTGATTGCTC AGGATACAGC 6420

TGAGAGAGTG AAACAGGCGG CGCATGAACG TTCAAACAAT ATCATTCATC AAGCAGAGCA 6480

AGATGCGCAA CGCTTGTTGG AAGAAGCTAA ATATAAGGCA AACGAGATTC TTCGTCAAGC 6540

AACTGATAAT GCTAAGAAAG TCGCTGTTGA AACAGAAGAA TTGAAGAACA AGAGCCGTGT 6600

CTTCCACCAA CGTCTCAAAT CTACAATTGA GAGTCAGTTG GCTATTGTTG AATCTTCAGA 6660

TTGGGAAGAT ATTCTCCGTC CAACAGCTAC TTATCTTCAA ACCAGTGATG AAGCCTTTAA 6720

AGAAGTGGTT AGCGAAGTAC TTGGAGAACC GATTCCAGCT CCAATTGAAG AAGAACCAAT 6780

TGATATGACA CGTCAGTTCT CTCAAGCAGA AATGGCAGAA TTACAAGCTC GTATTGAGGT 6840

AGCCGATAAA GAATTGTCTG AATTTGAAGC TCAGATTAAA CAGGAAGTGG AAGCTCCAAC 6900

TCCTGTAGTG AGTCCTCAAG TTGAAGAAGA GCCTCTGCTC ATCCAGTTGG CCCAATGTAT 6960

GAAGAACCAG AAGTAGCTCC AATGCATCCG ATAGGTCCAA CACCAGCTAC AGAAACTGTT 7020

GATTCAATAC CGGGATTTGA AGCACCGCAA GAATCTGTTA CAATTTTATA AGAAATATTC 7080

TGAGAACAAT ATCTTATCCT TATATTTCCA GCGAGCAGGA GATGGTGTGA GTCCTGTAAT 7140

CCCTATTGAT AAGATTATCC TCTCAAAAAC TCAAGTCTGA AGCTAGTAAG ATTTGACGTT 7200

TCCCACGTTA CGGGATAAGA GGGAGAAAGA CTAAATCTTT TTCCGAATAA AGGTGGTACC 7260

ACGATTTTCG TCCTTTTTGG AAGTCGTGGT TTTTAATTTG TTATTATTTA TAAAGGAGAT 7320

ACCATGAAAC TCAAAGACAC CCTTAATCTT GGGAAAACTG AATTCCCAAT GCGTGCAGGC 7380

CTTCCTACCA AAGAGCCAGT TTGGCAAAAG GAATGGGAAG ATGCAAAACT TTATCAACGT 7440

CGTCAAGAAT TGAACCAAGG AAAACCTCAT TTCACCTTGC ATGATGGCCC TCCATACGCT 7500

AACGGAAATA TCCACGTTGG ACATGCTATG AACAAGATTT CAAAAGATAT CATTGTTCGT 7560

TCTAAGTCTA TGTCAGGATT TTACGCACCA TTTATTCCTG GTTGGGATAC TCATGGTCTG 7620

CCAATCGAGC AAGTCTTGTC AAAACAAGGT GTCAAACGTA AAGAAATGGA CTTGGTTGAG 7680

TACTTGAAAC TTTGCCGTGA GTACGCTCTT TCTCAAGTAG ATAAACAACG TGAAGATTTT 7740

AAACGTTTGG GTGTTTCTGG TGACTGGGAA AATCCATATG TGACCTTGAC TCCTGACTAT 7800

GAAGCAGCTC AAATTCGTGT ATTTGGTGAG ATGGCTAATA AGGGTTATAT CTACCGTGGT 7860

GCTAAGCCAG TTTACTGGTC ATGGTCATCT GAGTCAGCAC TTGCTGAAGC AGAGATTGAA 7920

TACCATGACT TGGTTTCAAC TTCCCTTTAC TATGCCAACA AGGTAAAAGA TGGCAAAGGA 7980

GTTCTAGATA CAGATACTTA TATCGTTGTC TGGACAACGA CTCCATTTAC CATCACAGCT 8040

TCTCGTGGTT TGACGGTTGG TGCAGATATT GATTACGTTT TGGTTCAACC TGCTGGTGAA 8100

GCTCGTAAGT TTGTCGTTGC TGCTGAATTA TTGACTAGCT TGTCTGAGAA ATTTGGCTGG 8160

GCTGATGTTC AAGTTTTGGA AACTTACCGT GGCCAAGAAC TCAACCACAT CGTAACAGAA 8220

CACCCATGGG ATACAGCTGT AGAAGAGTTG GTAATTCTTG GTGACCACGT TACGACTGAC 8280

TCTGGTACAG GTATTGTCCA TACAGCCCCT GGTTTTGGTG AGGACGATTA CAATGTTGGT 8340

ATTGCTAATA ATCTTGAAGT CGCAGTGACT GTTGATGAAC GTGGTATCAT GATGAAGAAT 8400

GCTGGTCCTG AATTTGAAGG TCAATTCTAT GAAAAGGTAG TTCCAACTGT TATTGAAAAA 8460

GTTGGTAACC TCGTTCTTGC CCAAGAAGAA ATCTCTCACT CATATCCATT TGACTGGCGT 8520

ACTAAGAAAC CAATCATCTG GCGTGCAGTT CCACAATGGT TTGCCTCAGT TTCTAAATTC 8580

CGTCAAGAAA TCTTGGACGA AATTGAAAAA GTGAAATTCC ACTCAGAATG GGGTAAAGTC 8640

CGTCTTTACA ATATGATCCG TGACCGTGGT GACTGGGTTA TCTCTCGTCA ACGTGCTTGG 8700

GGTGTTCCAC TTCCTATCTT CTACGCTGAA GATGGTACAG CTATCATGGT AGCTGAAACT 8760

ATTGAACACG TAGCTCAACT TTTTGAAGAA TATGGTTCAA GCATTTGGTG GGAACGTGAT 8820

GCCAAAGACC TCTTGCCAGA AGGATTTACT CATCCAGGTT CACCAAACGG CGAGTTCAAA 8880

AAAGAAACTG ATATCATGGA CGTTTGGTTT GACTCAGGTT CATCATGGAA TGGAGTGGTG 8940

GTAAACCGTC CTGAATTGAC TTACCCAGCC GACCTTTACC TAGAAGGTTC TGACCAATAC 9000

CGTGGTTGGT TTAACTCATC ACTTATCACA TCTGTTGCCA ACCATGGCGT AGCACCTTAC 9060

AAACAAATCT TGTCACAAGG TTTTGCCCTT GATGGTAAAG GTGAGAAGAT GTCTAAATCT 9120

CTTGGAAATA CTATTGCTCC AAGCGATGTT GAAAAACAAT TCGGTGCTGA AATCTTGCGT 9180

CTCTGGGTAA CAAGTGTTGA CTCAAGCAAT GACGTGCGTA TCTCTATGGA TATCTTGAGC 9240

CAAGTTTCTG AAACTTACCG TAAGATTCGT AACACTCTTC GTTTCTTGAT TGCCAATACA 9300

TCTGACTTTA ACCCAGCTCA AGATACAGTC GCTTACGATG AGCTTCGTTC AGTTGATAAG 9360

TACATGACGA TTCGCTTTAA CCAGCTTGTC AAGACCATTC GTGATGCCTA TGCAGACTTT 9420

GAATTCTTGA CGATCTACAA GGCCTTGGTG AACTTTATCA ACGTTGACTT GTCAGCCTTC 9480

TACCTTGATT TTGCCAAAGA TGTTGTTTAC ATTGAAGGTG CCAAATCACT GGAACGCCGT 9540

CAAATGCAGA CTGTCTTCTA TGACATTCTT GTCAAAATCA CCAAACTCTT GACACCAATC 9600

CTTCCTCACA CTGCGGAAGA AATCTGGTCA TATCTTGAGT TTGAAACAGA AGACTTCGTC 9660

CAATTGTCAG AATTACCAGA AGTTCAAACT TTTGCTAACC AAGAAGAAAT CTTGGATACA 9720

TGGGCAGCCT TCATGGACTT TCGTGGACAA GCACAAAAAG CCTTGGAAGA AGCTCGTAAT 9780

GCAAAAGTTA TCGGTAAATC ACTTGAAGCA CACTTGACAG TTTATCCAAA TGAAGTTGTG 9840

AAAACTCTAC TCGAAGCAGT AAACAGCAAT GTAGCACAAC TTTTGATCGT GTCTGAGTTG 9900

ACCATCGCAG AAGGACCAGC TCCGGAAGCT GCCCTTAGCT TCGAAGATGT AGCCTTCACA 9960

GTTGAACGTG CTACTGGTGA AGTATGTGAC CGTTGCCGTC GTATCGACCC AACAACAGCA 10020

GAACGCAGCT ACCAGGCAGT TATCTGTGAC CACTGTGCAA GCATCGTAGA AGAAAACTTT 10080

GCGGAAGCAG TCGCAGAAGG ATTTGAAGAG AAATAAGATT GAAAAGTCTA GGCAAAATTC 10140

AATTTGAGAA GAAAAGACAA CTAATTTTAT AGTCTATTAA ACGCATTGTA TCACGTTTTT 10200

GAATACCTGA TATGATGCGT TTTTTATTTA TTTTAAAAAT TTGCGAGGTA TGACTTTTTA 10260

TACTCAACAA GAATCAAAGA GAAACTTAGC AAGCTAACAG TAGTAAGATA AAATAGGAAT 10320

TTGATATTAG GGATAAGATT GGTAAATAGT GTAATATTTT TACAACAATA AATTTATATA 10380

GTTATTTCTG GTTTCTGAAA AGTATTATAT TTTATTTCAT ATTATACAAA TTTTTATTTT 10440

ATAATATCAG AACATACTTT TTTTAAAAGC AAATATGATA CAATTTTATT TGAAAAAAAT 10500

AAAAAAGGAG ATTTTATTAT AAAATTAAAA AGACTTGCTT TAATTAGTGG TATCGTCGGT 10560

CTTGTGGGAG GAATTTTACT TCTTATTGGT CCTTTTGTCT TGTTGGGAAT AGCGGTAAAC 10620

ACAGCTGCTA CAACTCTTAA TGGAGGAGCT ACTGCAGGGG CTTTTTCAGG TGTAGCCTTA 10680

CTCTTGAATG CCTTGAAGAT TGCAAATCTT GTTCTTGGTA TCATTGCTAT TGTTTACTAT 10740

AAAGGAGATA AGCGTGTAGG TGCAGCTCCG TCTGTACTAA TGATTGTTTC TGGTGGAGTT 10800

AGTCTCATTC TATTCCGTTC TTAGGATGGG TTGGGGGGAT TTTTGCTATT ATCGGAGGAT 10860

CTCTATTCCT TTCAACATTG AAGAAATTCA AATCAGAAGA ATAAAAGGTA TTTTAGCATG 10920

AAAAGAACAA AAAAGTTTAT CGGTATAGGA GTAGCTCTAT TATCTCTTTC TCTTCTAGTT 10980

GCATGTGGAA CATAAAGTTC AAAGAATACT TCAACAAGTA ATGATGAGAA GACAGTAGCA 11040

ACATCCAATA GTTCAAAAGA AACAATCACT TTCGATACAC CGGTTGTAAC AGACGATGCG 11100

ATTGAATCAA TACGCACTTA TGCAGATTAT ATAGATCTTT ATAAAAATAT TTTTGATGAT 11160

TATTTTACTA AAGCTGAGGA AGGTTTCAAA GGCATAGCTA TGGAAAATAA TGACTCGTTT 11220

ACTAAACTAA AAGAGTCAAC TCAAAAATTA TTCGATGCGC AGAAAAAAAG GTTAAATAAT 11280

GAAGATAGAA TAGAAACAAC CAAAAACAAT GTGATTGCCA AACATTGTCA AACAGTCCTT 11340

TCCTTTTTGG TTTTGACTAG CTTTTTTGTG AAAAATTGTG TAAAATAGAA TAGATAAACG 11400

AGGGGAAACC TCGGAAAATT TAAAGGAGAA TCCATCTAAT GGTAAAATTG GTTTTTGCTC 11460

GCCACGGTGA GTCTGAATGG AACAAAGCTA ACCTTTTCAC TGGTTGGGCT GATGTTGATT 11520

TGTCTGAAAA AGGTACACAA CAAGCGATTG ACGCTGGTAA ATTGATCAAA GAAGCTGGTA 11580

TCGAATTTGA CCAAGCTTAC ACTTCAGTAT TGAAACGTGC TATCAAAACA ACTAACTTGG 11640

CTCTTGAAGC TTCTGACCAA TTGTGGGTTC CAGTTGAAAA ATCATGGCGC TTGAACGAAC 11700

GTCACTACGG TGGTTTGACT GGTAAAAACA AAGCTGAAGC TGCTGAACAA TTTGGTGATG 11760

AGCAAGTTCA CATCTGGCGT CGTTCATACG ATGTATTGCC TCCAAACATG GACCGTGATG 11820

ATGAGCACTC AGCTCACACA GACCGTCGTT ACGCTTCACT TGACGACTCA GTTATCCCAG 11880

ATGCTGAAAA CTTGAAAGTG ACTTTGGAAC GTGCTCTTCC ATTCTGGGAA GATAAAATCG 11940

CTCCAGCTCT TAAAGATGGT AAAAACGTAT TCGTAGGAGC TCACGGTAAC TCAATCCGTG 12000

CCCTTGTAAA ACACATCAAA GGTTTGTCAG ATGACGAGAT CATGGACGTG GAAATCCCTA 12060

ACTTCCCACC ATTGGTATTC GAATTCGACG AAAAATTGAA CGTCGTTTCT GAATACTACC 12120

TTGGAAAATA AAAAATTGTA AGTCTAGAAT TGATTTCTAG GCTTTTTATG TTAGTATGGA 12180

AGTATGATAA GGAATAAAAA ACAAGATTAT GTACTGGCCT ACAAGCAACC AGCTTCAACC 12240

ACTTACATGG GTTGGGAAGA AGAAGCTTTA CCGATAGGCA ATGGTTCTTT AGGAGCAAAA 12300

GTATTTGGCC TTATAGGGGC TGAACGGATT CAATTTAATG AAAAAAGTCT CTGGTCTGGA 12360

GGTCCACTTC CTGATAGTTC AGATTATCAG GGTGGAAATC TTCAGGATCA GTATGTTTTT 12420

TTAGCTGAGA TTCGGCAGGC TTTGGAGAAG AGAGATTACA ATCTGGCTAA GGAACTGGCT 12480

GAGCAGCACC TAATTGGGCC AAAAACGAGT CAATATGGGA CCTATCTGTC TTTTGGGGAT 12540

ATTCACATTG AGTTCAGCCA GCAAGGTACG ACTTTGTCTC AGGTGACGGA CTATCAGAGA 12600

CAGCTGAATA TTAGTAAGGC ACTTGCGACG ACTTCTTATG TCTATAAGGG AACGCGATTT 12660

GAACGTAAAG CTTTTGCGAG TTTTCCAGAT GATCTCTTGG TTCAATGTTT TACTAAGGAA 12720

GGGTTGGAAA CTCTAGATTT TACTATAGAA CTATCCTTGA CCTGTGATTT GGCTTCTGAT 12780

GGAAAGTATG AGCAGGAAAA ATCTGATTAC AAGGAGTGTA AGTTGGATAT TACTGATTCT 12840

CATATCTTGA TGAAGGGAAG AGTTAAGGAT AATGATCTGC GGTTTGCTAG TTATCTAGCT 12900

TGGGAAACGG ATGGAGATAT TAGAGTTTGG TCAGATAGGG TTCAGATATC AGGAGCCAGT 12960

TATGCCAATC TCTTCTTGGC CGCTAAGACG GATTTTGCCC AAAATCCTGC TAGCAATTAT 13020

CGCAAGAAAC TAGATTTAGA GCAACAGGTG ATAGACTTGG TGGACACAGC TAAAGAAAAG 13080

GGCTATACCC AATTGAAATC AAGGCATATC GAGGACTACC AAGCCTTATT CCAGCGTGTT 13140

CAATTGGATT TGGAAGCTGA TGTTGACGCA TCCACTACAG ATGATTTGTT AAAAAATTAT 13200

AAGCCACAAG AAGGGCAGGC TTTGGAGGAG CTGTTCTTCC AGTATGGACG GTATTTATTG 13260

ATTAGTTCGT CCAGAGACTG CCCAGATGCT CTACCAGCTA ACCTACAGGG AGTCTGGAAT 13320

GCGGTCGACA ATCCTCCTTG GAATTCGGAC TATCACTTAA ATGTCAATCT GCAGCTGAAT 13380

TATTGGCCAG CCTATGTTAC CAATCTCCTA GAGACGGTCT TTCCAGTCAT CAACTATGTA 13440

GATGATTTGC GTGTCTATGG TCGTCTAGCG GCTGTAAAGT ATGCAGGAAT CGTCTCTCAG 13500

AAAGGTGAGG AGAATGGTTG GTTGGTTCAT ACTCAAGCGA CTCCCTTTGG TTGGACGGCA 13560

CCTGGTTGGG ATTACTATTG GGGTTGGTCA CCAGCTGCCA ATGCGTGGAT GATGCAAACC 13620

GTTTATGAAG CCTATTTATT TTATAGGGAC CAAGACTATC TCAGGGAGAA AATTTATCCC 13680

ATGTTGAGGG AAACGGTTCG TTTTTGGAAT GCCTTTTTAC ATAAGGATCA GCAGGCGCAG 13740

CGTTGGGTGT CTTCTCCGTC TTATTCCCCA GAACATGGGC CGATTTCGAT TGGCAATACC 13800

TATGACCAAT CTCTGATTTG GCAGTTATTT CATGATTTTA TTCAGGCTGC TCAGGAATTG 13860

GGACTGGATG AGGACTTGTT GACTGAGGTT AAGGAGAAGT CTGATTTACT AAATCCTTTG 13920

CAAATCACTC AATCTGGTCG AATCAGGGAG TGGTATGAGG AGGAAGAGCA GTATTTTCAA 13980

AATGAGAAAG TGGAGGCCCA GCATCGGCAC GCTTCCCATC TAGTGGGACT CTATCCTGGC 14040

AATCTCTTTA GCTACAAGGG ACAAGAGTAT ATTGAAGCGG CGCGTGCTAG CCTCAATGAT 14100

CGTGGAGATG GCGGCACAGG CTGGTCCAAG GCTAATAAGA TCAATCTCTG GGCGCGTTTG 14160

GGAGATGGCA ATCGAGCCCA TAAATTATTG GCAGAGCAGT TAAAGACATC CACCTTGCAA 14220

AATCTTTGGT GTAGCCATCC TCCTTTTCAG ATAGATGGTA ATTTTGGTGC TACTAGTGGC 14280

ATGGCAGAAA TGTTACTCCA GTCTCATGCA GCTTATCTGG TACCTCTAGC TGCCCTACCT 14340

GATGCTTGGT CAACAGGTTC TGTTTCAGGC TTAATGGCAC GTGGACATTT TGAAGTGAGC 14400

ATGAGCTGGG AAGATAAAAA ACTCTTACAG TTGACCATTT TATCAAGGAG TGGAGGAGAT 14460

TTGCGAGTTT CTTATCCAGA TATTGAGAAG AGTGTGATTA AAATGAATCA AGAAAAAATA 14520

AAAGCGAAAT GCATGGGGAA AGATTGTATT TCGGTGGCAA CAGCAGAAGG TGATCTTGTT 14580

CAATTTTATT TTTAAGAAGA TGTTATAAGG CAGTAATTTG AAACTGCCTT TTAATAAGGA 14640

TTTAAGAATA TAAGCAGTTT TCAACTAGTT GAAAAAACGT TATAATGATA ATAGGAAGTA 14700

ATACTCAATG AAAATCAAAG AGCACAAACT AGGAAGCTAG CCGCAGGTTG CTCAAAACAG 14760

TGTTTTGAGG TTGCAGATGG AAGCTGACGT GGTTTGAAGA GAGATTTTCG AGGAGTATAA 14820

TTTGTTTGAT AGAGGGTGGG TCTGATGGCT TATATTGAGA TGAAACACTG TTACAAGCGT 14880

TATCAGGTTG GGGACACGGA GATTGTGGCC AATTGTGATG TGAATTTTGA GATTGAAAAG 14940

GGGGAGCTGG TTATTATCCT TGGTGCTTCA GGTGCAGGCA AGTCAACAGT TCTTAACCTT 15000

CTTGGGGGAA TGGATACCAA TGATGAAGGG GAAATCTGGA TTGATGGTGT TAATATTGCG 15060

GATTATAGTT CCCACCAGCG CACCAATTAC CGTAGAAATG ATGTGGGGTT TGTTTTTCAG 15120

TTTTATAATC TAGTTTCTAA TCTGACAGCT AAGGAAAATG TGGAACTGGC TTCTGAAATT 15180

GTGACAGATG CCTTGAATCC TGATCAGGCC TTGACAGATG TAGGTCTGGC TCATCGTCTC 15240

AATAACTTTC CAGCCCAGCT TTCTGGAGGG GAGCAACAGC GAGTCTCCAT TGCACGCGCG 15300

GTAGCCAAAA ATCCTAAAAT TCTCCTTTGT GATGAACCGA CTGGAGCCTT GGATTATCAG 15360

ACGGGCAAGC AGGTTTTGAA AATTCTCCAA GACATGTCTC GTCAAAAGGG AGCGACGGTC 15420

ATCATCGTGA CTCATAATGG AGCTTTGGCG CCCATTGCTG ATCGCGTGAT TCAAATGCAC 15480

GATGCCAGTG TCAAGGATGT GGTGCTCAAC CAGCATCCTC AGGATATTGA CAGTTTGGAG 15540

TACTAGCATG ATCAAGCGAA AAACTTATTG GAAGGACTTA GTTCAGTCCT TCACAGGCTC 15600

CAAGGGGCGT TTTTTATCCA TCTTGATCCT GATGATGTTG GGATCTCTAG CCTTAGTAGG 15660

CCTCAAAGTA ACCAGTCCCA ACATGGAGGC GACAGCTAAT GCTTATTTAA CAACTGCTCA 15720

AACCTTGGAT TTGGCAGTCA TGTCTAACTA TGGCTTGGAT CAAGCAGACC AAGAAGAACT 15780

AAAACAGACG GAGGGCGCAG AGGTCGAGTT TGGCTATTTG ACAGATGTGA CTATGGATAA 15840

TGGGCAGGAT GCCATTCGGC TGTACTCCAA ACCAGAGCGA ATTTCAACCT TTCAGCTAAG 15900

AAAGGGACGA CTTCCTCAGT CAGACAAGGA AATCGCTTTG GCCACTCATT TGCAAGGCCA 15960

ATACAGCGTG GGACAGGAGA TTAGTTTTAA AGAAAAAGAA GAGGGTCATT CCTCTTTAAA 16020

AGACCATACT TATACCATTA CTGGTTTTGT GGATTCGGCT GAAATCCTCT CCCAGCGAGA 16080

TATGGGCTAC GCAGGAAGTG GAAGTGGGAC TCTGACAGCC TATGGGGTGA TTTTACCTAG 16140

TCAATTTGAT CAGAAAGTCT ACAATATAGC TCGTTTGAAA TATCAAGATT TAGCGGGTTT 16200

AAATGCCTTT TCATCAGCTT ATGAAGAAAA ATCCAAGCAA CATCAAGAAG AGCTTGAACA 16260

AATTTTATCA GATAATGGCA AGGTACGTCT GCAACTTTTG AAAAAAGAAG GACAAGAGTC 16320

TCTAGACAAG GGGCAAGAGA CCCTTGACAA GGCTCAGACT AATTTGCAGG AAGGCAAGCG 16380

TCGTTTAGCA GCTGCTCAAG CTCGTATACA GGCTCAAGAA AGTCAACTAG CCTTGTTTCC 16440

TCAAGTTCAG AGAGAGCAGG CTAGTGCTCA ACTTACCCAA GCCAAGCAGG AATTGGGCAA 16500

GGAAGAGGAC AAACTAAAGC AAGCTGAACA AAATCTAGCC CAAGAAAAGG AAAAATTAGA 16560

AAAACATCAG CAAGTCTTGG ATGATTTGGC GGAGCCAAGG TATCAGGTTT ATAATCGTCA 16620

GACCATGCCA GGTGGTCAGG GCTATCTTAT GTATAGCAAT GCTTCATCCA GTATTCGAGC 16680

AGTGGGCAAT ATCTTTCCTG TGGTACTTTA TGCCGTAGCA GCCATGGTGA CCTTTACGAC 16740

CATGACTCGC TTTGTAGACG AAGAGCGAAC TCATGCAGGG ATTTTTAAGG CCTTGGGTTA 16800

TCGTAGTAAG GATATTATCG CCAAGTTTCT CCTTTATGGA CTAGTAGCTG GGACTGTCGG 16860

AACGGCTCTA GGTAGTATAC TTGGTCATTA TTTGCTAGCC AGTGTAATTT CAAGTGTCAT 16920

TACAAAAGGC ATGGTGGTGG GAGAAACTCA GATTCAGTTC TATTGGACCT ATAGCTTACT 16980

AGCTTTTGTC TTGAGCTTGT TGGCGAGTGT GTTACCAGCC TATCTGGTGG CTTGGAGGGA 17040

ACTTCATGAC GAAGCAGCCC AGCTTCTACT TCCTAAACCT CCTGTCAAAG GAGCTAAAAT 17100

CTTATTGGAG CGTATCGGTT TTATCTGGCG TCGTCTCAGT TTTACTCATA AGGTAACAGC 17160

CCGCAACATC TTTCGTTATA AGCAGAGAAT GTTGATGACA ATCTTTGGTG TGGCAGGTTC 17220

TGTAGCTCTG CTCTTTGCAG GTTTGGGAAT CCAATCTTCT GTAGCAGGAG TTCCGTCTAA 17280

ACAGTTTCAA CAAATCCAAC AGTATCAGAT GCTTGTCTCT GAAAATCCTA GTGCGACCAA 17340

TCAGGACAAG GTAGAGCTAG CAGAAGTGTT GAAAGGGCAG GAGATACTAG CCTACCAGAA 17400

AATCTATTCT AAAGCGCTAT ACAAGGATTT CAAAGGCAAA GCTGGTCTTC AAAACATTAC 17460

TCTTATGATG ATAGAGAAGG AAGATTTGAC TCCCTTTATC CATCTTCAAC ATCATCAGCA 17520

GGAGCTGACA TTAAAAGATG GCATCGTTAT TACAGCTAAA CTCGCCCAGC TGGCAGGTGT 17580

CAAGGTTGGG CAGACTTTAG AAATTGAAGG TAAGGAACTA AAGGTCGTTG CTATTACTGA 17640

GAACTACGTT GGTCACTTTA TTTATATGAG TCAGGCTAGC TATGAGCAAC TTTACGGACA 17700

GCTACCCCAA GCCAACACTT ATCTGGTCTC ATTAAGGGAT ACCAGTGCAA CTAGTATCGA 17760

AAGTCAGGCG GGCTTGCTTA TGAATCAATC TGCGGTGTCC AGCGTTGTCC AAAATGCTTC 17820

AGCCATTCGA CTCTTCGACT CTATCGCTAG CTCACTCAAT CAGACCATGA CCATCTTGGT 17880

CATCGTATCG GTTCTATTAG CTATTGTCAT CCTTTACAAT CTGACCAATA TCAACGTAGC 17940

TGAGAGAATC CGTGAACTCT CCACTATCAA GGTTCTTGGT TTTCATAATA ATGAAGTCAC 18000

CCTCTACATT TACCGTGAGA CGATTGTGCT GTCCCTTGTG GGAATCGTAC TTGGTCTGAT 18060

AGCTGGTTTC TATTTACACC AATTTTTGAT TCAAATGATT TCGCCTGCGA CTATTCTCTT 18120

TTATCCGCAG GTAGGCTGGG AAGTCTATGT AATCCCAGTG GCAGCAGTAA GCATCATTTT 18180

GACCTTGCTT GGTTTCTTCG TCAATTATTA TCTGAGAAAG GTTGATATGT TAGAAGCCCT 18240

GAAATCTGTA GAGTAAGGTA GTTATTTTTA GCTGATTGAA CTTCTATTTA CTAATATTCA 18300

AAAATCCTCC GTTTCAAAGA GCAGGGAACT CTTTGTGACA GAGGATTTTT TCTATAGGGC 18360

TTTAGCAGCT GCAATTGCGG CTTCGAAGTT TGGCTCAGAA TTGATATTAT CCACGTATTC 18420

AACGTAGCGA ATCGTATTGT CAGTATCGAG GACAAAGACT GCGCGTGCTA ATAGGTGCCA 18480

TTCGTTGATC AAGAGGGCAT AATCGCGCCC GAAAGAATGG TCAAAGTAGT CTGAAAGCAT 18540

AATGGCATTG TCAAGGCCTT CAGCACCGCA CCAACGTTTT TGAGCAAAAG GTAGGTCCAT 18600

TGAAACAGTC AATACGACCG TGTTGTCCAG TCCAGCCAAT TCTTCATTAA AACGACGTGT 18660

TTGAGTTGAG CAGATGCCTG TATCGATAGA AGGAACGACA CTCAAGACTT TTTTCTTGCC 18720

ATCAAAATCA GCCAGAGATT TTTTAGAAAG ATCTGTTGTA GTAAGAGAAA AATCAAGCGC 18780

CTTGTCGCCG ACTTGTAGTT GTTTACCTGT AAAGCTCACA GGATTTCCGA GAAAAGTTAC 18840

CATAGGATAC TCCAATCTTT TTTCTTCCAT TTTAGCTGAA ACAGTCGGAA TTTTCCAATG 18900

ATTTGACCGG AAATATGGGC ATAGAAAAAA CGCCAGCTCA TGTGAGAATG ACGTTTTTCA 18960

TAGGTTTATT TTGCCAATCC TTCAGCAATC TTGTCAAGGT TGTATTTCAT CATGCTGTAG 19020

TAGCTGTCGC CTTCTTTACC TTGTTCTGCG ATAGAGTCAG TAAAGATTTG AGCGTAGATT 19080

GGGATGTTTG TGTCTTGAGA AACAGTTTTC ATTGGACGGT CATCCACACT TGATTCTACA 19140

AAGAGTGATG GAACTTTTGT TTGGCGAAGT TTTTCAACCA AGGTCTTGAT TTGTTCAGGA 19200

GTTCCTTCTT CTTCAGTATT GATTTCCCAG ATGTAAGCAC TTGGGACACC ATAGGCTTTA 19260

GAGAAGTATT TGAATGCTCC TTCGCTGGTT ACAATGAGTT TCTTTTCAGC AGGGATCTTA 19320

TTAAATTTAT CCTTACTTTC TTTATCAAGT TTGTCTAACT TATCAGTATA TTCTTTGAGA 19380

TTTTTTTCAT AGAATTCTTT ATTGTTAGGG TCTTTGGCGC TCAATTGTTT GGCGATATTT 19440

TTAGCAAAAA TAATACCGTT TTCAAGGTTA AGCCAAGCGT GTGGGTCTTC TTTTCCTTTT 19500

TCATTTTGAC CTTCAAGGTA GATAACATCA ACGCCGTCGC TGACTGCGAA GTAGTCTTTG 19560

TTTTCAGTTT TCTTGGCATT TTCTACCAAT TTTGTAAACC AAGCATTGCC ACCTGTTTCA 19620

AGGTTGATAC CGTTATAGAA AATCAAATTA GCCTCAGAAG TTTTCTTAAC GTCTTCAGGA 19680

AGTGGTTCGT ATTCGTGTGG GTCTTGCCCA ATCGGAACGA TACTATGAAG GTCAATTTTG 19740

TCACCAGCAA TATTTTTAGT AATATCAGCG ATGATTGAGT TTGTAGCAAC AACTTTTAGT 19800

TTTTGACCAG AAGTTGTATC TTTTTTTCCG CTAGCACATG CTACAAGAAT GATTGCAGAA 19860

AGAAAGAGAA CGAGTAATGT ACCTAATTTT TTCATTAGAT CCTCCAATTT ATTAGGGCTT 19920

TGCCCCTTAT TTTAACAAAT GTTTATTTTT CAGTTTCAAA TATCGTTGTT TGGGAGCGAT 19980

AAAGAAGCTA ATGAGAAAGA AACTAGCAGC TGTAAGCACG ATACTAGAAC CTGCCGCAAC 20040

ATTAAAACTA TAGCCAATAA AGAGTCCCAA AACTGAAGCA GTAGCTCCGA AGGTTGAGGA 20100

AAGGAAAATC ATACTTTTCA GACTATTAGC ATACAGATAA GCAGTTGCAG CTGGGGTAAT 20160

CAGCATGGCT ACAATCAGGA TAGTTCCGAC ACTTTGCATG GCTGTCACAG ACACGAGAGT 20220

CAGGAGTACC ATGAGAAGGT AGTGATAGAA ATTGACAGGC ATTCCCATGG CTTTAGCCAA 20280

GAGTTCATCA AAGGAAGTTA TCAAGAGTTG CTTGAAGAAA ATCCAGATTA ACAAGAGGAT 20340

AGCTGCCCCC ACACCCATAG TAATAAACAT ATCCGTATCT TGGACGGCCA GGATATTACC 20400

AAAAAGGATA TGGAAAAGGT CAGTTGAACT TTTAGCGACA CCAATCAAGA TGATACCGAG 20460

GGCTAAGAAA GAAGAAAAGG TAATGCCGAT GGCGGTATCG CTTTTGATAA TCGAGTTTCC 20520

TTTGATGTAG GTAATGATGA TGGCAGCTAG CAATCCAAAG ACAATGGCTC CGATAAAGAA 20580

GTCAAGGCCC AAGATGAAGG ATAGGGCTAC ACCTGGTAAG ACAGCATGTG AAATGGCATC 20640

TCCCATGAGT GACATCCCGC GTAGAATAAT GAAACATCCC ACAGCTCCAG CTACAATCCC 20700

GACGACAATA GCTGTTATCA AGGCATTTTG TAGGAAATGG AATTTTTGCA ATCCATCGAT 20760

AAATTCTGCA ATCATAGGTC ACCTCCATTG AAAAAGAGTT GATTACCGTA AGCTTCTTTT 20820

AGATTGGTTT CGGTAAAAGT TTCTTTTGTT GGACCAAAGG CAATCACTTC TCGATTGACA 20880

AGTAAGACTT GATCGAAGTA GTGGGGAATC TTGCTGAGGT CGTGGTGAAC GATGAGAACC 20940

GTCTTCCCAG CTTTTTTCAA ATCTCTCAGC GTATTCATGA TGATTTCCTC ACTGACAGAG 21000

TCAATCCCAG CAAAGGGTTC ATCCAAGAGG ATATAGTCGG CTTCCTGCAC CAAACATCTG 21060

GCAATCAAGA CCCGCTGGAA TTGACCTCCA GACAGTTGAC TAATTTGACG TTCAGCGTAG 21120

TCAGCTAGGC CGACGATTTC AAGGGCCTCT TGCACTTTCT TCCAATGTTT AGCCTTTAAA 21180

CTTCGAAAGA GAGGAATAGA GGGAAATAGT CCTAACGAGA CGCATTCCTT GACCTTGATG 21240

GGAAAGTTGT AGTCGATATT GATTTTTTGT TCGACATAGG CAATTCGGTG TAAGGATTTT 21300

TTAACTTCCT TGTCATCGAG AAATGCCTGA CCTTGATGTG GGATAATTCC CAACATACCT 21360

TTTAATAGTG TTGATTTCCC AGCGCCGTTT GGACCAATGA TGCCGGTAAT TGTTGGTCCA 21420

TGGAGCACTA GTGAAATATC CTTAAGTGCC AACGTTTCTT TGTAGGAGAC ACTGAGGTTT 21480

TCGATACGTA TCATAAACTT GTATTCCTCC TGTCTCTTAA TATACATTAA AAAAAAAATT 21540

AAGTCAAGTT AATTTTTGAA AAAATTAAAA TAATAACTGA AAAATAGATT CTAAAGATAA 21600

CTTTCAGGAT AAATTTCTAA ATTATAAAAC GCATAGTATC AAGTGTAAAA AACTTGGAAT 21660

TATGCGTTTT ATCATGGAAA GATTTTTTAT AATAGCTAAA AAATAA 21706

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6171 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GATCCCCAGG AAAAACCGAG GTTTTCCCAA TCAATCGTTA CTGTCATATT CCACTCCTTA 60

TTCTAAAAAC CTATTTCTTA TATTCTACAC TATTTTTCTA AAATAGCAAG TATATTTTGT 120

AATTTTCAGA AAATTTCTCC AATAAAAACC AACTCTTAGA ACTGATTCTT CATTTCACTT 180

ATTTATCTTC AGTAACTACT TCCTGAAGAT AAGCGTCAAA AACTTCTTCA TCTGAAATCG 240

TGTCAGAAAT GAAGCTTCCA TTGCTAGTGC GTTCTGACAA GTTCAAGTCT TGCAATCGGC 300

TTTCATAGAT TGTTCCTTTA TTGGATTGGA CAAGCAGAGT TTGGTCGTTC ACATCCACTT 360

CCGTACTGAA GAAATCGCCA ACAAATCCTT GCTCTGCAAC TGCTCCTGCC AAGAAGACAC 420

GATGCGGTTT GTTTTTCAAC TCACGCAAGA CTTGTAATCC TCGTTTGGCA CGGCTGGTTG 480

CTAGAATTTC CTCAATGGAA ACACGTTTCA AGCTTCCACG CTGGGTCAAG AGGTAGAAGG 540

ACGAAGTATT ACAGATAAAG CCAGATTGGA GCACATCATC TTCTTTCAAA TTCATAGCCT 600

TGACACCTGC TGCCTTAGCA CCGACAACCG GAACCTCTTC GATATTGAAA CGCAGGGCAT 660

AACCATTTTG ACTAACCAAG ACAACATCAT CTAGTTTAAT CGGAGCCACT GCTACAATCT 720

GATCTGTATC GTCTTTGAGC TTAGCATACT TGACAGACTT AGATCTATAG GTCCGCCATG 780

GAGTGAATTC TTTTCGCTCT ACCCGTTTGA TTTGACCAAG GCGAGTCACT GCAAAGTAGG 840

TTGTCGCATC GTCAAACTGA TCCAGTACTT CCACATAAAG GATTTCTTCA TTCGTTTCAA 900

AGTTTGTGAT GGTTTGGCTC AGATGCTCTC CGATGTCCTT CCAACGAATA TCTGCCAACT 960

CATGGATTGG TCTGTAGATG ACATTTCCAA GACTTGTGAA CATCAAGAGG TGCTGGGTTG 1020

TCTTGGCAGA TTGAACAAAA ATCAAACGGT CATCATCACG CTTGCCAATT TCTTCCAAGG 1080

TGGAAGCCGC AAAGGAACGT GGACTGGTAC GCTTGATGTA ACCTGCCTTG GTCACGCTGA 1140

CGTAGGTATC TTCCTCAGCG ATAAGACTAG CTGTATCAAT CTCAATTGCT TTCGCAGTGT 1200

CTTCTAAAGA ACTCAAACGA GGAGTTGCAA ATTTCTTCTT GACCTCACGA AGTTCTTTCT 1260

TCATGAGATT GTACATAGTC CTTTCATCAC CGATAATAGC CGCCAGCATA GCAATCTTCT 1320

CACGAAGCTC TGCTTCTTCT TCCTGCAAGA CAACCACATC GGTATTGGTC AAACGGTACA 1380

GTTGCAAAGT TACGATAGCC TCAGCCTGTT CTTCCGTAAA ATCATAGCTA ACTTTGAGGT 1440

TTTCCTTGGC GTCCGCCTTA TTCTCAGAAG CACGGATAAG AGCAATGACT TCATCCAAAA 1500

TCGAAATCAC ACGAATCAAA CCTTCGACGA TATGGAGACG TTTCTCAGCC TTTTCTTTGT 1560

CAAAGCGTGA ACGCGCCAAA ATCACTTCTC GACGGTGAGC GATATAGCTA GACAGGATTG 1620

GAACAATCCC AACCTGACGA GGTGTGAAAT TGTCAATCGC CACCATATTA AAGTTGTAGT 1680

TGATTTGTAG GTCGGTGTAC TTAAATAAGT AGTTGAGAAC AAGCTCAGTA TTAGCGTCTT 1740

TCTTAAGTTC GATAGCGATA CGAAGACCAT CACGGTCAGA CTCATCACGA ACCTCAGCAA 1800

TCCCAGCTAC CTTGTTATTA ACACGAACAT CATCGATTTT CTTGACTAGA TTGGCCTTAT 1860

TGATTTCATA AGGAATCTCA ATAATAACGA TTTGTTCCTT ACCACCTTTT AGCTTTTCAA 1920

TTTCAGTCTT GGAACGAACA ACCACGCGCC CTTTCCCAGT CTCATAAGCT TTCTTGATTT 1980

CATCACGACC CTGAATAATA GCCCCTGTAG GGAAGTCTGG TCCAGGCAAG AATTCCATGA 2040

GTTTATCAAT CTTTGCAGTT GGGTGGTCAA TCATGTAAAC TGCAGCATCT ATGACCTCAG 2100

CTAAATTATG GGGAGGAATG TCTGTGGCAT AACCAGCCGA AATCCCAGTC GAACCATTGA 2160

CCAAGAGGTT TGGAAAGGCT GCTGGCAAGA CCGTTGGTTC TTTCTCCGTA TCGTCAAAGT 2220

TCCATGCAAA AGGAACTGTC TTTTTCTCGA TATCCTGAAG AAGGTAGCCT GCAATTTCAG 2280

ACAAACGTGC CTCAGTATAA CGCATAGCCG CAGGAGGATC TCCGTCCATA GAACCGTTAT 2340

TACCGTGCAT TTCAACTAGA ATCTCACGAT TTTTCCAGTT CTGTGACATA CGAACCATGG 2400

CATCATAGAT AGAAGAATCC CCGTGTGGGT GGAAATTCCC CATGATGTTC CCGACTGACT 2460

TGGCCGACTT ACGGTAGCTC TTGTCAAAAG TATTGCTATC CTTATTCATA GAATAAAGAA 2520

TACGGCGCTG AACCGGCTTC AACCCATCAC GAATATCTGG CAAAGCCCGG TCTTGAATAA 2580

TGTACTTGGA GTAGCGACCA AAGCGCTCTC CCATGATGTC CTCCAGGGAC ATGTTTTGAA 2640

TGTTAGACAT AAGATACAAA GCCCATAAAA TACCAAGTGA AAATAGAAAA TTCTTGAAGT 2700

AAGCAAACTC ACAAGAGAAT TTATCTTTTT CACACAGTAT CTAGGGCGTG TTCAACTCCT 2760

TTCAAAGAAT GTAGAGTAGG TTTTTATGCA GTAAAAGATA TTTTACGGGA ATTCCTCCCG 2820

TGTTCAGTTA CGATAAGTAA CCAAACTATC CTGTTTGTAT TTTTCAATAT GAAAATCTGG 2880

TTTTCCAAAA TTAGTCTTAG TTTGTGTCTT AGCCGCTCCC TTAAGCGCCT CTTTGAGATA 2940

AGCACTCATA GCAGATTCTT CATTAATAAT CCTGCAATTT TTTCAAACCA AGATTTTCAA 3000

ACTGCTTTTT CACATAGTCA TTCACATCCG ACTCTAATTT CCAGTTTACT AACATATTAT 3060

TTTCTTTCAT TAAAACACTG TCGTTTCTTC TAGCGTAAAC TTGACATTAT CTTCAATCCA 3120

TTTACGGCGT GGTTCTACCT TATCTCCCAT GAGAACATTG ACGCGGCGTT CGGCGCGCGC 3180

TAAATCTTCA ATTGTGACAC GGATGAGGGT ACGTGTTTCT GGGTTCATGG TTGTTTCCCA 3240

GAGCTGGTCC GCATTCATCT CACCAAGTCC TTTGTATCGT TGGAGGGTAG CGCCTTTACC 3300

GAACTGTTTA CGGAGTTCTT CTAGTTCTCC GTCCGTCCAA GCGTAGGCCA CTTCTTCTTT 3360

CTTGCCTTTA CCTTTGGACA TCTTGTAAAG AGGTGGGAGG GCAATATAGA CATGACCTGC 3420

CTCGACTAGC GGACGCATGT AACGGTAGAA AAATGTCAAG AGCAAGGTCT GGATATGGGC 3480

ACCGTCGGTA TCCGCATCGG TCATGATAAT GATCTTATCA TAGTTGGCAT CTTCAATAGA 3540

GAAGTCTGCT CCAACACCCG CACCAATGGT ATAAATCATG GTATTGATCT CTTCATTTTT 3600

GAGGATATCC GCCATCTTGG CCTTGGCTGT ATTGACAACC TTACCACGAA GAGGTAGAAT 3660

AGCCTGGAAC TTGCGGTCAC GACCTTGTTT GGCAGAACCA CCGGCAGAGT CCCCCTCAAC 3720

TAGATAGAGT TCATTCTTAG CAGGATTCTT AGATTGGGCT GGGGTCAATT TCCCAGACAA 3780

CAAGCCCTTA TCTTTCTTGT TTTTCTTCCC ATTTCGGCTC TCATCACGCG CCTTACGTGC 3840

TGCTTCACGA GCATCACGGG CCTTGATAGC CTTGCGGATG AGGTTAGAAG CTAATTCCCC 3900

ATTTTCCATA AGGAAAAAGG TCAACTTATC AGCCACTATT CCATCCACAA CTGGGCGAGC 3960

TAGGGGGCTT CCTAGTTTAT CCTTGGTCTG TCCTTCAAAC TGCAAGTGTT CTTCAGGAAC 4020

TAAGATAGAA AGAACGGCCG CTAGTCCCTC ACGATAGTCT GAACCTTCAA GGTTTTTATC 4080

TTTTTCCTTG AGAAGACCTG TTTTACGTGC ATAGTCATTC ATGACCTTGG TAATGGCAGA 4140

CTTGAGTCCT GTCTCGTGCG TTCCACCGTC CTTGGTGCGA ACGTTATTGA CAAAAGATAG 4200

AATGTTATCT GAGAATCCGT CATTGTACTG GAGGGCTACT TCCACTTGAA AACCATTGTC 4260

TTCCCCTTCA AAGTAAAGAA CTGGCGTCAA GATTTCCTTA TCTTCGTTGA GATAAGAAAC 4320

AAAATCTTGT ACTCCATTCT CATAGTGGAA CTCAATCGCT TCATTTGTTC GCTTGTCCGT 4380

TAAAGACAAG GTCACATTTT TCAAGAGAAA GGCTGATTCA TTAAGGCGCT CTGAAATGGT 4440

ATTGTACTTG AAATCTGTCG TAGAAAATAT AGTCGCGTCA GGCATAAAAG TAACTTTGGT 4500

GCCTGTTTTA GACTTGGGTG CTGTACCGAT TTTCTTCAAA GTCGTGACAG GTTTTCCACC 4560

ATTTTCGAAA CGTTGCTTGT AAACTGCGCC ATCACGGGTA ATTTCAACTT CTAACCAGCT 4620

AGAAAGGGCG TTAACAACGG AAGAACCCAC TCCGTGAAGT CCACCTGATG TCTTATAGCC 4680

ACCTTGACCG AATTTCCCTC CGGCATGAAG AATGGTAAAG ATAACCTCAA CAGTTGGAAT 4740

TCCCATAGCG TGCATACCTG TCGGCATCCC ACGTCCATGG TCTTGAACCG TTAGACTACC 4800

GTCTTTATTG ATAGTTACAT CAATACGATC ACCAAACCCA GACAAGGCTT CATCGACTGC 4860

ATTATCAACG ATTTCCCAAA CTAGGTGATG AAGACCAGCG CCATCGGTCG ATCCAATATA 4920

CATCCCTGGA CGTTTTCGGA CCGCATCCAA CCCTTCTAGC ACCTGAATAG CATCATCATT 4980

ATAATTGTTA ATATTGATTT CCTTTTTTGA CACAAGGAAC CTCCTATTCG TTCATCTTTA 5040

CTATTCTACA GGTTTTCCAA GGATTTTGCA AAATTTTTCT TTCTCCGATG TGACAATTTC 5100

AGCAGAGATT CTCTGCTTTT CTTTCCCAAT TCATGATATA ATAGGAGTAT GATTACAATA 5160

GTTTTATTAA TCCTAGCCTA TCTGCTGGGT TCGATTCCAT CTGGTCTCTG GATTGGACAA 5220

GTATTCTTTC AAATCAATCT ACGCGAGCAT GGTTCTGGTA ACACTGGAAC GACCAACACC 5280

TTCCGCATTT TAGGTAAGAA AGCTGGTATG GCAACCTTTG TGATTGACTT TTTCAAAGGA 5340

ACCCTAGCAA CGCTGCTTCC GATTATTTTT CATCTACAAG GCGTTTCTCC TCTCATCTTT 5400

GGACTTTTGG CTGTTATCGG CCATACCTTC CCTATCTTTG CAGGATTTAA AGGTGGTAAG 5460

GCTGTCGCAA CCAGTGCTGG AGTGATTTTC GGATTTGCGC CTATCTTCTG TCTCTACCTT 5520

GCGATTATCT TCTTTGGAGC TCTCTATCTT GGCAGTATGA TTTCACTGTC TAGTGTCACA 5580

GCATCGATTG CGGCTGTTAT CGGGGTTCTG CTCTTTCCAC TTTTTGGTTT TATCCTGAGT 5640

AACTATGACT CTCTCTTCAT CGCTATTATC TTAGCACTTG CTAGTTTGAT TATCATTCGT 5700

CATAAGGACA ATATAGCTCG TATCAAAAAT AAAACTGAAA ATTTGGTCCC TTGGGGATTG 5760

AACCTAACCC ATCAAGATCC TAAAAAATAA AATGCCAGTT CTGTACTGCC CCCAAACAGT 5820

TAGACAAATA ATTTATCCAA AGGATTTAGT TCTGTACTGC ACAGGACTAA GTCCTTTTAG 5880

TTTTACCTTA ATTCGTTTGT TGTTGTAGTA ATCAATATAG TCTATAATGG CTTGTTCCAA 5940

TTGATTAAGT GATTTAAATG TTTTCTCATA GCCATAAAAC ATTTCGGATT TTAAAATGCC 6000

AAAGAAAGAT TCCATCCTAC CGTTGTCTTG GCTGTTGCCC TTACGTGACA TGGATGCTTG 6060

AATTCCCTTA CTCTCTAGGA ACCGATGATA AGAATCGTGT TGGTATTGCC AGCCTTGGTC 6120

ACTATGGAGA ATCGTATTCT CGTAGTGCTT CTCTGTGAAT GCCTGTTCCA A 6171

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18475 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

TATTACAAAT AAAAAAACGG AGGAGTGCTT TATGAAAGCC TATACTTATG TTAAACCAGG 60

ACTTGCTTCT TTTGTTGATG TAGACAAACC AGTTATTCGC AAGCCAACAG ACGCTATTGT 120

GCGTATTGTA AAAACCACTA TTTGTGGAAC AGACCTCCAT ATTATCAAAG GGGATGTTCC 180

TACTTGCCAA AGTGGTACCA TTCTTGGCCA CGAAGGGATT GGGATTGTTG AAGAAGTTGG 240

GGAAGGAGTT TCCAACTTCA AAAAAGGTGA CAAGGTCTTG ATTTCTTGCG TCTGTGCCTG 300

TGGTAAATGC TACTACTGTA AAAAAGGAAT TTATGCTCAC TGTGAAGACG AAGGGGGCTG 360

GATTTTCGGT CACTTGATTG ATGGTATGCA GGCTGAATAT CTACGTGTCC CTCATGCAGA 420

TAATACTCTT TACCATACTC CAGAAGACTT GTCAGATGAA GCTTTGGTTA TGCTGTCAGA 480

CATTCTGCCT ACTGGATATG AAATTGGTGT CTTAAAAGGG AAAGTAGAAC CTGGTTGCAG 540

CGTAGCCATT ATTGGTTCAG GTCCAGTTGG ATTGGCTGCT CTTTTAACAG CCCAATTCTA 600

TTCACCAGCT AAATTGATTA TGGTAGACCT AGACGATAAC CGCTTGGAAA CTGCCCTATC 660

ATTCGGTGCG ACTCATAAGG TTAATTCTTC AGACCCTGAA AAAGCCATTA AAGAAATTTA 720

TGATTTGACA GATGGTCGTG GTGTGGATGT CGCTATCGAA GCTGTTGGTA TTCCTGCAAC 780

ATTTGATTTC TGTCAAAAGA TTATCGGTGT AGACGGAACG GTTGCCAACT GTGGTGTGCA 840

TGGTAAACCA GTTGAATTCG ATTTAGATAA ACTTTGGATT CGCAACATCA ATGTAACAAC 900

TGGTTTGGTA TCTACAAATA CGACTCCACA ATTGTTGAAA GCACTTGAAA GTCATAAGAT 960

TGAACCGGAA AAATTGGTAA CTCACTATTT CAAACTCAGT GAAATTGAAA AAGCCTACGA 1020

AGTCTTCAGT AAGGCAGCAG ACCACCATGC CATTAAGGTC ATTATCGAAA ACGATATCTC 1080

AGAAGCCTAA GTAGTAAAAA TATTTTTGTA CATAAGTAAA TAGAAATTCA GTCATCCATC 1140

AGATGGCTGG ATTTTTTATC AAAAAATTAA GAAATGAGCA TATTTCTTTC CTTGTCTGGC 1200

GGAATTGGTT ATAATATACG GTACAAAGGA ATGAATGAAT ATGTATCGTG TTATAGAAAT 1260

GTACGGAGAT TTTGAACCGT GGTGGTTCTT AGAAGGTTGG GAAGAAGATA TTGTAGCAAG 1320

TAGAAAATTT GACCAGTATT ATGATGCTCT CAAATACTAC AAAACTTGCT GGTTTAGATT 1380

GGAACAAGAA TCGCCTCTTT ATAAAAGTAG AAGCGACTTG ATGACCATTT TTTGGGACCC 1440

GGAAGACCAA CGCTGGTGTG ATGAATGTGA TGAGTATTTA CAACAATACC ATTCTTTGGC 1500

TCTTTTGCAG GATGAGCAGG TTATCCCAGA CGAAAAACTA CGCTCAGGCT ATGAAAAACA 1560

AACCAGTCAG GAAAGGAATC GTTCTTGCCG TATGAAATTA AAATAGAGAA AAGTAACTTT 1620

TTTGGAGTTG CTTTTTTTAT TTTTCTAACT CTTTGCGAAT AGTATAGGTG AGGAGGTAAG 1680

TATGGTTCAA GAAATTGCAC AAGAAATCAT TCGTTCAGCT CGGAAAAAAG GGACGCAGGA 1740

TATCTATTTT GTCCCTAAGT TAGACGCCTA TGAGCTTCAT ATGAGGGTAG GAGACGAGCG 1800

CTGTAAAATT GGTAGCTATG ATTTTGAAAA GTTTGCAGCC GTTATCAGTC ACTTTAAGTT 1860

TGTGGCGGGT ATGAATGTGG GAGAAAAAAG ACGTAGTCAA CTGGGTTCCT GTGATTATGC 1920

CTATGACCAT AAGATAGCGT CTCTACGTTT ATCTACTGTA GGCGATTATC GGGGGCATGA 1980

GAGTTTGGTT ATCCGTTTGT TGCACGATGA GGAGCAGGAC CTGCATTTTT GGTTTCAGGA 2040

TATTGAAGAA TTAGGCAAGC AGTACAGGCA ACGGGGACTC TATCTTTTTG CTGGTCCGGT 2100

TGGGAGTGGT AAGACGACCT TGATGCATGA ATTGTCCAAG TCACTCTTTA AAGGACAGCA 2160

AGTTATGTCC ATCGAAGATC CTGTCGAAAT CAAGCAGGAC GACATGCTTC AGTTGCAGTT 2220

GAACGAAGCA ATCGGCCTAA CCTATGAAAA TCTAATCAAA CTTTCCTTGC GTCATCGACC 2280

AGATCTCTTG ATTATCGGAG AAATTCGTGA CAGCGAGACG GCGCGTGCAG TGGTCAGAGC 2340

TAGTTTGACA GGTGCGACAG TCTTTTCAAC CATTCACGCC AAGAGTATCC GAGGTGTTTA 2400

TGAGCGTCTG CTGGAGTTGG GTGTGAGTGA AGAAGAATTG GCAGTTGTTC TGCAAGGAGT 2460

CTGCTACCAG AGATTAATCG GGGGAGGAGG AATCGTTGAC TTTGCAAGCA GAGATTATCA 2520

AGAACACCAA GCAGCCAAGT GGAATGAGCA AATTGACCAG CTTCTTAAAG ATGGACATAT 2580

CACAAGTCTT CAGGCTGAGA CGGAAAAAAT TAGCTACAGC TAAGCAAAAA AATATCATCA 2640

CCCTATTTAA CAATCTCTTT TCTAGCGGTT TTCATCTGGT GGAGACTATC TCCTTTTTAG 2700

ATAGGAGTGC TTTGTTGGAC AAGCAGTGTG TGACCCAGAT GCGTGTGGGC TTGTCTCAGG 2760

GGAAATCATT CTCAGAAATG ATGGAAAGTT TGGGATGTTC AAGTGCTATT GTCACTCAGT 2820

TATCCCTAGC TGAAGTTCAT GGCAATCTCC ACCTGAGTTT GGGAAAGATA GAAGAATATC 2880

TGGACAATCT GGCTAAGGTC AAGAAAAAAT TGATTGAAGT AGCGACCTAT CCCTTGATTT 2940

TGCTGGGTTT TCTTCTCTTA ATTATGCTGG GGCTACGGAA TTACCTGCTC CCACAACTGG 3000

ATAGTAGCAA TATTGCCACC CAAATTATCG GTAATCTGCC CCAAATTTTT CTAGGCATGG 3060

TAGGGCTTGT TTCCGTGCTT GCCCTTTTAG CACTCACTTT TTATAAAAGA AGTTCTAAGA 3120

TGAGTGTCTT TTCTATCTTA GCACGCCTTC CCTTTATTGG AATCTTTGTG CAGACCTACT 3180

TGACAGCCTA TTATGCACGT GAATGGGGGA ATATGATTTC ACAGGGAATG GAGTTGACGC 3240

AGATTTTTCA AATGATGCAG GAACAAGGTT CCCAGCTCTT TAAAGAAGTC GGTCAAGATC 3300

TGGCTCAAAC CCTGAAAAAT GGCCGTGAAT TTTCTCAGAC GATAGGAACC TATCCTTTCT 3360

TTAGGAAGGA ATTGAGTCTC ATCATAGAGT ATGGGGAAGT TAAGTCCAAG CTGGGTAGTG 3420

AGTTGGAAAT CTATGCTGAA AAAACTTGGG AAGCCTTTTT TACCCGAGTC AACCGCACCA 3480

TGAATTTGGT GCAGCCACTG GTTTTTATCT TTGTGGCACT GATTATCGTT TTACTTTATG 3540

CGGCAATGCT CATGCCCATG TATCAAAATA TGGAGGTAAA TTTTTAAAAT GAAAAAAATG 3600

ATGACATTCT TGAAAAAAGC TAAGGTTAAA GCTTTTACAT TGGTGGAGAT GTTGGTGGTC 3660

TTGCTGATTA TCAGCGTGCT TTTCTTGCTC TTTGTACCTA ATCTGACCAA GCAAAAAGAA 3720

GCAGTCAATG ACAAAGGAAA AGCAGCTGTT GTTAAGGTGG TGGAAAGCCA GGCAGAACTT 3780

TATAGCTTAG AAAAGAATGA AGATGCTAGC CTAAGAAAGT TACAAGCAGA TGGACGCATC 3840

ACGGAAGAAC AGGCTAAAGC TTATAAAGAA TACAATGATA AAAATGGAGG AGCAAATCGT 3900

AAAGTCAATG ATTAAGGCCT TTACCATGCT GGAAAGTCTC TTGGTTTTGG GACTTGTGAG 3960

TATCCTTGCC TTGGGCTTGT CCGGCTCTGT CCAGTCCACT TTTTCAGCGG TAGAGGAACA 4020

GATTTTCTTT ATGGAGTTTG AAGAACTCTA TCGGGAAACC CAAAAACGCA GTGTAGCCAG 4080

TCAGCAAAAG ACTAGTCTGA ACTTAGATGG GCAGACGCTT AGCAATGGCA GTCAAAAGTT 4140

GCCAGTCCCT AAAGGAATTC AGGCCCCATC AGGCCAAAGT ATTACATTTG ACCGAGCTGG 4200

GGGCAATTCG TCCCTGGCTA AGGTTGAATT TCAGACCAGT AAAGGAGCGA TTCGCTATCA 4260

ATTATATCTA GGAAATGGAA AAATTAAACG CATTAAGGAA ACAAAAAAATT AGGGCAGTGA 4320

TTTTACTGGA AGCAGTAGTC GCTCTAGCTA TCTTTGCCAG CATTGCGACC CTCCTTTTGG 4380

GACAAATTCA AAAAAATAGG CAAGAGGAAG CAAAAATCTT GCAAAAGGAA GAAGTCTTGA 4440

GGGTAGCTAA GATGGCCCTG CAGACGGGGC AAAATCAGGT AAGCATCAAC GGAGTTGAGA 4500

TTCAGGTATT TTCTAGTGAA AAAGGATTGG AGGTCTACCA TGGTTCAGAA CAGTTGTTGG 4560

CAATCAAAGA GCCATAAGGT CAAGGCTTTT ACCTTGTTAG AATCCCTGCT TGCCCTCATT 4620

GTCATCAGTG GGGGATTACT CCTTTTTCAA GCTATGAGTC AGCTCCTCAT TTCAGAAGTT 4680

CGCTACCAGC AACAAAGCGA GCAAAAGGAG TGGCTCTTGT TTGTGGACCA ACTTGAGGTA 4740

GAATTAGACC GTTCGCAGTT CGAAAAAGTA GAAGGCAATC GCCTATACAT GAAGCAAGAT 4800

GGCAAGGACA TCGCCATCGG TAAGTCAAAG TCAGATGATT TCCGTAAAAC GAATGCTCGT 4860

GGTCGAGGTT ATCAGCCTAT GGTTTATGGA CTCAAATCTG TACGGATTAC AGAGGACAAT 4920

CAACTGGTTC GCTTTCATTT CCAGTTCCAA AAAGGCTTAG AAAGGGAGTT CATCTATCGT 4980

GTGGAAAAAG AAAAAAGTTA AGGCAGGTGT TCTCCTCTAC GCAGTCACCA TAGCAGCCAT 5040

CTTTAGTCTT TTGTTGCAAT TTTATTTGAA CCGACAAGTC GCCCACTATC AAGACTATGC 5100

TTTGAATAAA GAAAAATTGG TTGCTTTTGC TATGGCTAAA CGAACCAAAG ATAAGGTTGA 5160

GCAAGAAAGT GGGGAACAGT TTTTTAATCT AGGTCAGGTA AGCTATCAAA ACAAGAAAAC 5220

TGGCTTAGTG ACGAGGGTTC GTACGGATAA GAGCCAATAT GAGTTTCTGT TTCCTTCAGT 5280

CAAAATCAAA GAAGAGAAAA GAGATAAAAA GGAAGAGGTA GCGACCGATT CAAGCGAAAA 5340

AGTGGAGAAG AAAAAATCAG AAGAGAAGCC TGAAAAGAAA GAGAATTCAT AGTCAATTCA 5400

ACTATAATGC GTTGAATCCA GAATAGTCCA CTGTAGTTTC TAGAAAATTG CTGGAAATGG 5460

ATGTTAAGCT CCAATTCATT TGTTTATATC TTATTTCAGT TTACTATACT TTGTGCTAAA 5520

TTAAAGATAT GAAACATGAT TTTAACCACA AAGCAGAAAC TTTCGATTCC CCTAAAAATA 5580

TCTTCCTCGC AAACTTGGTA TGTCAAGCAG CCGAGAAACA GATTGATCTT CTATCAGACA 5640

AAGAAATTTT AGATTTCGGT GGTGGCACGG GTCTATTAGC CTTGCCCCTA ACCCCTAGCC 5700

AAGCAGGCTA AGTCAGTCAC TCTTGTAGAC ATTTCTGAGA AAATGTTGGA GCAAGCTCGT 5760

TTGAAAGTGG AGCAGCAAGC AATCAAGAAT ATCCAGTTTT TGGAGCAAGA TTTACCGAAA 5820

AATCCCTTGG AGAAAGAGTT TGATTGCCTT GCTGTTAGTC GGGTTCTTCA TCATATGCCT 5880

GATTTGGATG CGGCTCTCTC ACTGTTTCAT CAACATTTGA AGGAAGATGG GAAACTCATC 5940

ATTGCTGATT TTACCAAGAC AGAAGCTAAT CATCATGGAT TTGATTTAGC TGAACTGGAA 6000

AACAAGCTAA TTGAGCATGG TTTTTCATCT GTGCATAGTC AGATTCTCTA TAGTGCTGAA 6060

GACCTGTTTC AAGGAAATCA CTCAGAATTC TTTTTAATAG TAGCCCAAAA ATCACTCGCC 6120

TAGTCAGGGA GTGATTTTTC TATAAGGATG GAAAAAAGAA GGGAAATTTG GTAAGATAGG 6180

AATATGGATT TTGAAAAAAT TGAACAAGCT TATACCTATT TACTAGAGAA TGTCCAAGTC 6240

ATCCAAAGTG ATTTGGCGAC CAACTTTTAT GACGCCTTGG TGGAGCAAAA TAGCATCTAT 6300

CTGGATGGTG AAACTGAGCT AAACCAGGTC AAGGAGAACA ATCAAACCCT TAAGCGTTTA 6360

GCACTACGCA AAGAAGAATG GCTCAAGACC TACCAGTTTC TCTTGATGAA GGCTGGGCAA 6420

ACAGAACCCT TGCAGGCCAA TCACCAGTTT ACACCGGATG CTATTGCTTT GCTTTTGGTG 6480

TTTATTGTGG AAGAGTTGTT TAAAGAGGAG GAAATTACTA TCCTCGAAAT GGGTTCTGGG 6540

ATGGGAATTC TAGGCGCTAT TTTCTTGACC TCGCTTACTA AAAAGGTGGA TTACTTGGGA 6600

ATGGAAGTGG ATGATTTGCT GATTGATCTG GCAGCTAGCA TGGCAGATGT AATTGGTTTG 6660

CAGGCTGGCT TTGTCCAAGG AGATGCCGTT CGCCCACAAA TGCTCAAAGA AAGCGATGTG 6720

GTCATCAGTG ACTTGCCTGT CGGCTATTAT CCTGATGATG CCGTTGCGTC GCGCCATCAA 6780

GTTGCTTCTA GCCAAGAACA TACTTACGCC CATCACTTGC TCATGGAACA AGGGCTTAAG 6840

TACCTCAAGT CAGACGGATA CGCTATTTTT CTAGCTCCGA GTGATTTGTT GACCAGTCCT 6900

CAAAGTGATT TGTTAAAAGA ATGGCTGAAA GAAGAGGCGA GTCTGGTTGC TATGATTAGT 6960

CTGCCTGAAA ATCTCTTTGC TAATGCCAAA CAATCTAAGA CTATTTTTAT CTTACAGAAG 7020

AAAAATGAAA TAGCAGTAGA GCCTTTTGTT TATCCACTTG CTAGCTTGCA AGATGCAAGT 7080

GTTTTAATGA AATTTAAAGA AAATTTTCAA AAATGGACTC AAGGTACTGA AATATAAAAT 7140

AGATTTTGTT ATAATAGTTG AAAACGCTTA AAAAGGGGTA TCATGTTATG ACAAAAACAA 7200

TTGCAATCAA TGCAGGAAGT TCAAGTTTGA AATGGCAATT ATACTTAATG CCAGAAGAAA 7260

AAGTATTGGC GAAAGGTTTG ATTGAACGTA TCGGTTTGAA AGATTCAATT TCAACTGTAA 7320

AATTTGACGG CCGTTCTGAA CAACAAATTT TGGATATTGA AAATCATATA CAAGCCGTTA 7380

AAATTTTATT GGATGACTTG ATTCGTTTCG ATATTATCAA GGCTTATGAC GAGATTACAG 7440

GTGTTGGACA TCGTGTTGTT GCTGGTGGAG AATATTTCAA AGAATCAACA GTTGTTGAGG 7500

GAGATGTTTT AGAAAAAGTT GAAGAGTTGA GTTTGTTGGC TCCTCTACAC AACCCGGCCA 7560

ATGCAGCAGG TGTTCGTGCC TTCAAGGAAT TGTTGCCAGA CATTACCAGT GTAGTTGTTT 7620

TTGATACTTC CTTCCACACA AGTATGCCAG AGAAAGCTTA TCGCTACCCT CTACCAACAA 7680

AATATTACAC AGAAAACAAG GTTCGTAAAT ACGGTGCTCA TGGTACAAGT CACCAGTTTG 7740

TAGCAGGAGA AGCTGCAAAA CTCTTGGGAC GTCCATTAGA AGACTTGAAG TTAATTACCT 7800

GTCATATTGG TAACGGAGGC TCAATTACAG CTGTGAAAGC CGGCAAATCT GTAGACACTT 7860

CTATGGGGTT CACTCCTCTT GGTGGTATTA TGATGGGAAC GCGTACAGGG GATATTGATC 7920

CAGCTATCAT TCCTTATTTA ATGCAATATA CAGAGGATTT TAACACACCA GAAGATATCA 7980

GTCGTGTTCT TAACCGTGAA TCAGGTCTTT TGGGAGTTTC TGCTAATTCT AGCGATATGC 8040

GCGATATAGA AGCAGCTGTA GCAGAAGGGA ATCACGAGGC TAGCTTGGCT TATGAAATGT 8100

ATGTTGACCG TATCCAAAAA CATATCGGTC AGTACCTTGC AGTGCTAAAT GGAGCAGATG 8160

CCATTGTTTT CACAGCAGGT GTCGGTGAAA ATGCAGAGAG TTTCCGTCGT GATGTAATCT 8220

CAGGGATTTC GTGGTTTGGT TGTGATGTTG ATGATGAAAA GAATGTCTTT GGCGTTACAG 8280

GAGACATCTC AACAGAGGCA GCTAAAATCC GTGTCTTGGT TATTCCAACA GATGAAGAAT 8340

TAGTCATTGC CCGTGACGTT GAACGCTTGA AAAAATAAGT GAAACTAAAA AAATATTCAA 8400

TACAAGGAGT TGGGAAAGTT ATTTTTCCAG CTTCTTTTTC TGATGAAATT GTCCAAAACC 8460

TTGCTATGAT TGGCTTTTTT GAAAAATATG GTATAATAGT AGTAATTTAA TAGATGGAGT 8520

TGAGTTTTGA AGAAAAACTT TCGTGTAAAA AGAGAGAAAG ATTTTAAGGC GATTTTCAAG 8580

GAGGGGACAA GTTTTGCTAA TCGCAAATTT GTGGTCTACC AATTAGAAAA CCAGAAAAAC 8640

CGTTTTCGAG TAGGTCTATC AGTTAGCAAA AAACTGGGGA ATGCCGTCAC TAGAAATCAA 8700

ATTAAGCGAC GGATTCGGCA TATTATCCAG AATGCAAAAG GGAGTCTGGT AGAAGATGTC 8760

380

GACTTTGTTG TCATTGCTCG AAAAGGAGTC GAAACCTTGG GATACCTAGA GATGGAGAAA 8820

AATCTACTCC ATGTATTAAA ATTATCAAAG ATTTACCGGG AAGGAAATGG GAGTGAAAAA 8880

GAAACTAAAG TTGACTAGTT TGCTAGGACT GTCTCTGTTA ATCATGACAG CCTGTGCGAC 8940

TAATGGGGTA ACTAGCGATA TTACAGCCGA ATCGGCTGAT TTTTGGAGTA AATTGGTTTA 9000

CTTCTTTGCG GAAATCATTC GCTTTTTATC GTTTGATATT AGTATCGGAG TGGGGATTAT 9060

TCTCTTTACG GTCTTGATTC GTACAGTCCT CTTGCCAGTC TTTCAGGTGC AAATGGTGGC 9120

TTCTAGGAAA ATGCAGGAAG CTCAGCCACG CATTAAGGCG CTTCGAGAAC AATATCCAGG 9180

TCGAGATATG GAAAGCAGAA CCAAACTAGA GCAGGAAATG CGTAAAGTAT TTAAAGAAAT 9240

GGGTGTCAGA CAGTCAGACT CTCTTTGGCC GATTTTGATT CAGATGCCGG TTATTTTGGC 9300

CCTGTTCCAA GCCCTATCAA GAGTTGACTT TTTAAAGACA GGTCATTTCT TATGGATTAA 9360

CCTTGGTAGT GTGGATACAA CCCTTGTTCT TCCGATTTTA GCAGCAGTAT TCACCTTTTT 9420

AAGTACTTGG TTGTCCAACA AAGCTTTGTC TGAGCGAAAT GGCGCTACGA CTGCGATGAT 9480

GTATGGGATT CCAGTCTTGA TTTTTATCTT TGCAGTTTAT GCGCCAGGTG GAGTCGCCCT 9540

ATACTGGACA GTGTCTAATG CTTATCAAGT CTTGCAAACC TATTTCTTGA ATAATCCATT 9600

CAAGATTATC GCAGAGCGCG AGGCCGTAGT ACAGGCACAA AAAGATTTGG AAAATAGAAA 9660

AAGAAAAGCC AAGAAAAAGG CTCAGAAAAC GAAATAAATA AGGAGGAATC TGGTAGTGGT 9720

AGTATTTACA GGTTCAACTG TTGAAGAAGC AATCCAGAAA GGATTGAAAG AATTAGATAT 9780

TCCAAGAATG AAGGCTCATA TCAAAGTCAT TTCTAGGGAG AAAAAAGGCT TTCTTGGTCT 9840

ATTTGGTAAA AAACCAGCCC AAGTGGATAT TGAAGCGATT AGTGAAACGA CTGTTGTCAA 9900

AGCAAATCAA CAGGTAGTAA AAGGCGTTCC GAAAAAAATC AATGATTTGA ACGAGCCTGT 9960

GAAGACGGTT AGTGAAGAAA CCGTTGACCT TGGTCATGTG GTTGATGCTA TTAAAAAAAT 10020

AGAGGAAGAA GGTCAAGGTA TTTCTGATGA AGTCAAGGCT GAAATCTTAA AACATGAAAG 10080

ACATGCCAGC ACTATCTTAG AAGAAACTGG TCACATTGAG ATTTTAAATG AACTTCAAAT 10140

CGAGGAAGCG ATGAGGGAAG AAGCAGGCGC TGATGACCTT GAAACTGAGC AAGACCAAGC 10200

TGAAAGTCAA GAACTAGAAG ACTTGGGCTT GAAAGTTGAA ACGAACTTTG ATATTGAACA 10260

AGTAGCTACG GAAGTAATGG CTTATGTTCA AACGATTATT GATGACATGG ATGTTGAGGC 10320

TACACTTTCA AATGATTATA ACCGTCGTAG CATCAATCTA CAAATTGACA CCAACGAACC 10380

AGGTCGTATT ATCGGCTACC ATGGTAAAGT CTTGAAGGCC TTGCAACTGT TGGCTCAAAA 10440

TTATCTTTAC AACCGCTATT CCAGAACCTT CTACGTTACA ATCAATGTCA ATGATTATGT 10500

CGAACACCGT GCAGAAGTCT TGCAGACCTA TGCGCAAAAA TTGGCGACTC GTGTTTTGGA 10560

AGAAGGGCGC AGTCATAAAA CAGATCCAAT GTCAAATAGC GAACGCAAGA TTATCCATCG 10620

TATTATTTCA CGTATGGATG GCGTGACTAG TTACTCTGAA GGTGATGAGC CAAATCGCTA 10680

TGTTGTTGTA GATACAGAAT AAGTAAAATC AGGTTTATCC TGATTTTTTG CTAGTTAGAG 10740

GAGGTTAAAC TGATGTTGAA TAAGATAAGA GACTATTTAG ACTTTGCTGG TTTGCAGTAC 10800

CGTAATCCTG ATAAAGCGGG AGCAGAGCGA GAGAAGATGC TGGCATTCCG CCACAAAGGA 10860

CAAGAGGCCC GAAAGGTTTT TACAGAACTG GCCAAAGCCT TTCAAGCAAG CCATCCAGAA 10920

TGGCAACTCC AACAGACTAG CCAGTGGATG AATCAGGCCC AGCGTTTGAG ACCACATTTT 10980

TGGGTTTATC TACAGAGAGA CGGACAAGTG ACAGAACCTA TGATGGCCTT ACGTTTGTAT 11040

GGGACATCTA CTGACTTTGG AATTTCTTTG GAAGTCAGTT TCATCGAACG TAAGAAGGAT 11100

GAGCAAACAC TGGGCAAGCA GGCCAAAGTT TTAGACATTC CAACCGTTAA AGGGATTTAT 11160

TATCTAACCT ACTCTAATGG TCAAAGTCAA CGGTGGGAGG CGAATGAAGA AAAGCGTCGT 11220

ACTTTACGCG AGAAGGTGAG AAGTCAAGAA GTTCGAAAAG TTTTAGTGAA GGTAGATGTT 11280

CCTATGACAG AAAATTCGTC TGAAGAAGAA ATCGTAGAAG GCTTATTGAA GTCTTATTCT 11340

AAAATTCTTC CCTATTATCT AGCTACGAGA AAATAAGATA ATTTGTAAAA CATCATAAAT 11400

CATACAGTCC AAGAGTGAAC AGTCCGCTGT GTAATTCTTG GTCTTTTTGT TTGCGCTTTC 11460

GCATTATATA ATAAACTTAC AAAAACAATT CAAAAGGAGA ACAATTATGG AAGTCGTTTC 11520

AAGTGTTCTA AATTGGTTTT CTAGCAATAT TTTGCAGAAT CCCGCATTTT TCGTAGGTTT 11580

ATTGGTGTTG ATAGGATATG CACTTTTGAA AAAACCTGCC CATGACGTTT TTTCAGGGTT 11640

TGTTAAAGCA ACAGTAGGGT ATATGTTGCT TAACGTGGGT GCTGGTGGTT TGGTTACAAC 11700

CTTTCGTCCA ATCTTAGCAG CTCTTAACTA CAAATTCCAA ATTGGTGCAG CGGTTATCGA 11760

CCCTTACTTT GGACTTGCTG CAGCAAACAA CAAAATTGTA GCAGAGTTTC CAGATTTTGT 11820

TGGAACTGCA ACTACAGCTC TATTGATTGG TTTTGGAATA AATATCTTGC TCGTAGCTCT 11880

TCGAAAGATT ACGAAGGTAA GAACCCTCTT TATTACTGGT CACATCATGG TACAACAAGC 11940

TGCAACAGTA TCTCTTATGG TTCTATTCTT AGTACCACAA TTGCGCAATG CTTACGGTAC 12000

AGCAGCGATT GGTATCATCT GTGGACTTTA CTGGGCAGTT AGTTCAAATA TGACTGTTGA 12060

GGCAACTCAA CGCTTGACTG GTGGTGGCGG ATTTGCGATT GGTCACCAAC AGCAATTTGC 12120

AATCTGGTTT GTAGATAAAG TAGCAGGACG CTTTGGTAAG AAAGAAGAAA GTTTAGACAA 12180

TCTTAAATTA CCTAAGTTCC TCTCAATCTT CCACGATACA GTTGTTGCAT CTGCTACCTT 12240

GATGCTCGTA TTCTTCGGAG CCATTCTTTT AATCTTGGGT CCAGACATTA TGTCTAATAA 12300

AGAAGTCATC ACTTCAGGAA CTCTATTCAA TCCTGCTAAA CAAGATTTCT TTATGTACAT 12360

TATCCAAACA GCCTTTACCT TCTCAGTTTA CTTGTTCGTT TTGATGCAAG GTGTCCGAAT 12420

GTTCGTATCT GAGTTGACAA ACGCCTTCCA AGGTATTTCA AACAAATTGT TGCCAGGTTC 12480

ATTCCCAGCG GTTGACGTTG CAGCTTCTTA TGGATTTGGT TCTCCAAATG CTGTCTTGTC 12540

AGGATTTACC TTTGGTTTGA TTGGTCAATT GATTACAATT GTTTTGCTCA TCGTCTTTAA 12600

AAATCCGATT CTTATTATTA CAGGATTTGT ACCAGTGTTC TTTGACAATG CAGCCATTGC 12660

GGTCTACGCT GATAAACGCG GCGGATGGAA AGCGGCTGTT ATCCTTTCCT TTATATCAGG 12720

TGTCCTTCAA GTTGCTCTAG GAGCTCTTTG TGTGGCCCTT CTCGATTTGG CATCTTATGG 12780

TGGCTACCAT GGAAATATCG ACTTTGAATT CCCATGGCTT GGATTTGGAT ATATCTTCAA 12840

ATACCTTGGT ATTGTTGGTT ATGTACTTGT GTGTCTCTTC TTGCTTGTTA TTCCTCAACT 12900

TCAATTTGCC AAAGCAAAAG ATAAAGAGAA ATATTACAAC GGTGAAGTTC AAGAAGAAGC 12960

TTAGTATCTA GAAAAGGAGA AATAAAATGG TTAAAGTATT AGCAGCGTGC GGAAATGGAA 13020

TGGGTTCATC AATGGTTATC AAGATGAAGG TTGAAAATGC TCTCCGTAAG CTTAATCAAA 13080

CAGATTTTAC AGTCAATTCA TGCAGTGTCG GTGAAGCTAA AGGTTTAGCA GTAGGATATG 13140

ACATCGTAAT CGCTTCTCTT CATTTGATTC AAGAATTGGA AGGGCGAACT AATGGGAAGT 13200

TAATTGGGCT TGATAACTTG ATGGATGATA AAGAAATCAC CGAAAAACTC AGTCAAGCAC 13260

TACAGTAAAA GGTTGGAGGG GGCTGGACAG AAACTGAGAG TTATCGTTTC TGTCCTTCTC 13320

CCTCTTTAAA TAAAGGAGGC AGATATGAAT TTAAAACAAG CTTTAATTGA CAATGACTCG 13380

ATCCGACTAG GTTTAGAGGC TAACAATTGG AAAGAAGCAG TCAAGGTAGC AGTAGATCCC 13440

TTAATTGAAA GTGGGGCAAT TTTGCCAGAG TATTACGATG CTATCATTGA ATCGACTGAA 13500

GAGTATGGGC CTTACTATAT CTTGATGCCA GGTATGGCTA TGCCCCACGC TAGACCTGAA 13560

GCAGGTGTGC AAAGTGATGC CTTTTCATTG ATTACCTTAC AAAATCCTGT TGTATTTTCA 13620

GATGGGAAAG AGGTATCTGT TTTGTTGGCA CTAGCAGCAA CAAGTTCAAA AATTCACACA 13680

AGTGTAGCCA TTCCACAAAT TATTGCCCTA TTTGAATTAG AAGATTCTAT TGCACGTTTA 13740

CAGGCTTGCC AGACTAAAGA AGATGTCTTG GCTATGATTG AAGAATCTAA GGATAGCCCT 13800

TATCTCGAAG GATTGGATTT GGAAAGTTAG AAAGAGGAAT AAAGAAATGA CAAAAAGAAT 13860

ACCTAATTTA CAAGTTGCAT TAGACCATTC AGACTTGCAA GGAGCGATTA AAGCAGCTGT 13920

TTCTGTTGGT CAGGAAGTAG ATATTATCGA AGCTGGAACT GTTTGCTTGC TTCAAGTTGG 13980

AAGTGAACTG GCTGAAGTCT TGCGTAGCCT TTTCCCAGAT AAGATTATTG TGGCAGACAC 14040

AAAATGTGCT GATGCTGGTG GAACAGTTGC TAAAAATAAT GCGGTTCGTG GAGCAGACTG 14100

GATGACTTGT ATCTGTTGTG CAACCATCCC TACTATGGAA GCAGCTCTAA AGGCTATCAA 14160

GACTGAACGA GGAGAACGAG GCGAAATCCA GATCGAGCTT TATGGCGATT GGACTTTTGA 14220

ACAAGCTCAG CTTTGGCTAG ATGCAGGTAT CTCACAAGCT ATTTATCACC AATCTCGTGA 14280

TGCTCTTCTT GCTGGTGAAA CTTGGGGTGA AAAAGACCTT AATAAGGTTA AAAAACTCAT 14340

TGACATGGGC TTCCGTGTAT CTGTAACAGG TGGTCTAGAT GTAGATACTC TCAAACTCTT 14400

TGAAGGTATT GATGTCTTTA CCTTTATCGC AGGTCGTGGA ATTACAGAGG CTGTGGATCC 14460

AGCAGGAGCA GCGCGTGCCT TCAAGGATGA AATCAAACGA ATTTGGGGGT AAATCATGGT 14520

ACGTCCAATT GGAATTTATG AAAAGGCAAC CCCAACACAC TGTACTTGGC TAGAACGTTT 14580

AAATTTTGCC AAGGAGTTAG GCTTTGATTT TGTCGAGATG TCTATTGACG AACGTGACGA 14640

GCGTTTAGCA AGACTTGACT GGAGTAAGGA AGAACGCTTG GAAGTTGTCA AAGCAATCTA 14700

TGAAACTGGT GTTCGTATTC CTTCTATCTG TTTTTCAGGC CATCGTCGCT ACCCATTGGG 14760

TTCAAAAGAT CCAGTTCTAG AGGAAAAATC TCTAGAACTC ATGAAAAAAT GTATCGAATT 14820

AGCTCAAGAC TTGGGAGTTC GTACGATTCA ATTAGCTGGT TACGATGTTT ACTATGAGGA 14880

AAAGTCACCC CAGACACGCC AACGTTTTAT CAAAAATTTG AGAAAAGCCT GTGACTGGGC 14940

TGAAGAAGCT CAGGTGGTAC TTGCTATTGA AATTATGGAT GATCCTTTCA TCAGTAGCAT 15000

CGAAAAATAT TTGGCTATAG AAAAAGAGAT TGACTCTCCC TTCCTCTTTG TATATCCAGA 15060

TATTGGTAAT GTGTCTGCAT GGCATAATGA TATCTATAGT GAGTTTTATC TTGGTCATCA 15120

TGCCATCGCA GCTCTCCATC TCAAGGATAC TTATGCAGTG ACAGAAAGTT CAAAGGGCCA 15180

GTTCCGAGAT GTACCTTTCG GGCAAGGTTG TGTCAAATGG GAAGAAGCTT TCGATATTTT 15240

AAAGGAAACC AATTATAATG GACCTTTCCT AATCGAAATG TGGTCTGAAA ATTGTGAAAC 15300

AGTAGAAGAA ACACGCGCAG CCATTCAAGA GGCGCAAGCT TTTCTCTATC CACTCATTAA 15360

GAAAGCAGGT TTGATGTAAG ATGAATCAAG TAATCAATGC TATGCGTAAA CGAGTCTGTG 15420

ATGCCAATCA ATCATTGCCA AAACATGGAC TTGTCAAATT TACCTGGGGG AATGTATCTG 15480

AAGTTAATCG CGAACTCGGT GTCATTGTTA TCAAACCATC AGGCGTGGAT TATGACGAAT 15540

TGACACCTGA AAACATGGTA GTGACTGATC TAGATGGTAA GATCCTAGAA GGGGATTTAA 15600

GACCATCTTC CGACCTCCCA ACTCATGTGC AATTATATAA GACTTGGTCA GAAATTGGTA 15660

GTGTGGTTCA CACCCATTCG ACAGAAGCTG TTGGTTGGGC TCAGGCAGGT CGTGATATTC 15720

CTTTCTACGG AACAACCCAT GCAGATTATT TCTACGGTTC AATCCCTTGC GCCCGTAGTT 15780

TGACCAAGGA CGAAGTAGAA GTGGCCTATG AAAAAGATAC TGGCCTGGTT ATCGTAGAAG 15840

AGTTTGAACA TCGCGGACTT AACCCGGTTG AAGTACCAGG AATTGTTGTA CGCAATCACG 15900

GTCCATTCAC CTGGGGCAAA AATCCAGAGA ATGCTGTTTA TCACTCTGTC GTACTAGAGG 15960

AAGTATCAAA GATGAATCGC TTTACAGAAC AAATCAATCC AAGAGTTGGA CCTGCTCCCC 16020

AGTACATACT AGAAAAACAC TACCAACGTA AACATGGACC AAATGCTTAT TATGGTCAAA 16080

AGTAAGAACG ATGAAGGAGG AGAAAAAGAT AAATTTAGCT CCTCTTTTTA CATTTGATTT 16140

TTATTGAGAG TAAAGTTGGA GTTGAAGTAA TTTTAAAAGA TTTTTTAGAA ATAGCGCTTG 16200

ATATATATAT GGTAAAATAA AAAGAATTGC TGTGATATCA ATAGATTTGG GGGATTTTTT 16260

AATATGGTAC TGGATAAGGC AAGTTGTGAT TTGCTTCAAT ATTTGATGGA TCAAGAAACG 16320

TCCAAAACGA TTATGGCGAT TTCGAAAGAT TTGAAAGAGT CAAGAAGGAA AATTTATTAT 16380

CACATTGACA AAATCAATGC TGCTCTGGGT GACGAGGCGC TTCACATCAT TAGTATTCCA 16440

CGAATTGGTA TTCACTTAAC GGAAGAGCAG AGAGATGCTT GTTGTAAACT ATTATCGGAA 16500

GTAGATTCGT ACGATTATAT CATGAGTGCG CATGAACGTA TGATGATAAT GTTACTATGG 16560

ATAGGTATTT CTAAAGAACG TATTACGATT GAAAAATTGA TAGAGTTAAC AGAGGTATCT 16620

AGGAATACTG TTCTCAATGA TTTGAATAGT ATTCGTTATC AACTAACTTT GGAACAATAT 16680

CAGGTGATCT TGCAAGTGAG CAAGTCACAG GGATACAACC TTCATGCCCA CCCTCTTAAT 16740

AAAATTCAGT ATCTTCAATC GCTTCTATAT CATATTTTTA TGGAAGAAAA TGCCACTTTT 16800

GTATCTATTT TAGAAGATAA GATGAAAGAG AGGTTAGATG ATGAGTGTTT GCTTTCTGTT 16860

GAAATGAACC AATTTTTTAA GGAACAGGTT CCTTTAGTTG AACAAGATTT AGGGAAGAAA 16920

ATAAACCATC ATGAAATAAC TTTTATGTTG CAGGTTCTAC CTTATTTGCT GTTAAGCTGT 16980

CATAATGTTG AACAGTATCA AGAAAGACAT CAGGATATAG AGAAAGAATT TTCTTTGATA 17040

AGAAAAAGAA TAGAGTATCA GGTGTCTAAG AAATTAGGAG AACGGTTGTT TCAAAAGTTT 17100

GAAATTTCTT TGTCAGGACT TGAAGTTTCT CTTGTAGCTG TTCTCCTCCT CTCCTATCGT 17160

AAAGATTTGG ATATTCATGC AGAAAGTGAT GATTTTCGGC AATTAAAACT TGCTTTAGAA 17220

GAATTTATCT GGTATTTTGA ATCACAAATC CGAATGGAGA TTGAGAACAA GGATGATTTG 17280

TTACGAAATT TGATGATCCA CTGTAAAGCC TTGTTATTTA GAAAGACTTA CGGTATTTTT 17340

TCTAAAAATC CTCTAACAAA ACAAATTCGA TCCAAGTATG GAGAATTATT TTTAGTCACT 17400

AGAAAATCTG CGGAAATTTT AGAAGGAGCA TGGTTTATTC GGCTAACAGA CGATGATATT 17460

GCCTATTTGA CGATTCATAT TGGAGGATTT TTAAAATATA CACCATCATC TCAAAAAAAT 17520

ATGAAAAAAG TTTATCTCGT TTGTGATGAA GGTGTTGCGG TTTCGAGACT TTTGCTGAAA 17580

CAATGCAAAC TTTATTTTCC AAATGAGCAA ATTGACACTG TATTTACAAC AGAACAATTT 17640

AAGAGTGTGG AAGATATTGC ACAAGTTGAT GTAGTGATTA CTACTAATGA TGATTTGGAT 17700

AGCAGATTTC CGATTTTAAG GGTTAATCCT ATCCTTGAAG CAGAAGATAT TTTGAAAATG 17760

CTAGACTATC TTAAACACAA TATATTTCGT AATAAGAGCA AAAGTTTCAG TGAAAATCTT 17820

TCTAGTCTTA TTTCGTCTTA TATTGTAGAC AGCAAGTTGG CTAGTAAGTT CCAAGAAGAG 17880

GTTCAAACAC TTATAAATCA AGAAATAGTA GTTCAAGCTT TTTTGGAAGr TATTTGAAGG 17940

ACAGTCCAAT GATGAACACA AACCTGTGTk TTTCsTGGTC TTTTtTAGTG TTTTGAAGGG 18000

TGGkATACTA ATCTCAAAGA TAACAATTAT ATCCAAAGGA GGCAACATAT GCCAAACGTC 18060

AAAGAAATTA CAAGAGAGTC ATGGATTTTA GCCACTTTCC CAGAGTGGGG AACATGGTTG 18120

AACGAAGAAA TCGAAGAAGA AGTCGTACCT GAAGGCAACT TTGCCATGTG GTGGCTAGGC 18180

AACTGTGGTA CTTGGATTAA GACACCAGCT GGTGCTAACG TTGTCATGGA CCTTTGGTCA 18240

AACCGTGGAA AATCAACCAA AAAAGTGAAA GATATGGTTC GTGGGCACCA AATGGCAAAT 18300

ATGGCAGGTG TTCGTAAGCT GCAACCAAAC TTGCGTGTTC AGCCAATGGT TATCGATCCA 18360

TTTGCTATCA ACGAACTAGA CTATTACTTA GTTTCACACT TCCACAGTGA TCATATCGAC 18420

CCATACACAG CTGCAGCAAT TCTCAATAAT CCTAAGTTAG AGCATGTTAA GTTGG 18475

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7186 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CCAGGATTTG GTACCGTTGC AAGTGGTGTG CCTTTCCTCC TAAAGGAAAA TGGAGGAAAA 60

ATCAATCAAT CAGCACATTC AGATATCAAA GTTGCTAAGG TATTGGTCAA GGATGAAGAT 120

GAAAAAAATC GCTTGCTTGC AGCAGGGAAT GACTTTAACT TTGTAACCAA TGTGGATGAT 180

ATTTTATCAG ACCAGGATAT TACTATCGTA GTGGAATTGA TGGGGCGTAT TGAGCCTGCT 240

AAAACCTTTA TCACTCGTGC CTTGGAAGCT GGAAAACACG TTGTTACTGC TAACAAGGAC 300

CTTTTAGCTG TCCATGGCGC AGAATTGCTA GAAATCGCTC AAGCTAACAA GGTAGCACTT 360

TACTACGAAG CAGCAGTTGC TGGTGGGATT CCAATTCTTC GTACTTTAGC AAATTCCTTG 420

GCTTCTGATA AAATTACGCG CGTGCTTGGA GTAGTCAACG GAACTTCCAA CTTCATGGTG 480

ACCAAGATGG TGGAAGAAGG CTGGTCTTAC GATGATGCTC TTGCGGAAGC ACAACGTCTA 540

GGATTTGCAG AAAGCGATCC GACGAATGAC GTAGATGGGA TTGATGCAGC CTACAAGATG 600

GTTATTTTGA GCCAATTTGC CTTTGGCATG AAGATTGCCT TTGATGATGT AGCCCACAAG 660

GGAATCCGCA ATATCACACC AGAAGACGTA GCTGTAGCTC AAGAGCTTGG TTACGTAGTG 720

AAATTGGTTG GTTCTATTGA GGAAACTTCT TCAGGTATTG CTGCAGAAGT GACTCCAACC 780

TTCCTACCTA AAGCGCACCC ACTTGCTAGT GTGAATGGCG TAATGAACGC TGTCTTTGTA 840

GAATCTATCG GTATTGGTGA GTCTATGTAC TACGGACCAG GTGCGGGTCA AAAACCAACT 900

GCAACAAGTG TTGTAGCTGA TATTGTCCGT ATCGTTCGTC GTTTGAATGA TGGTACTATT 960

GGCAAAGACT TCAACGAATA TAGCCGTGAC TTGGTCTTGG CAAATCCTGA AGATGTCAAA 1020

GCAAACTACT ATTTCTCAAT CTTGGCTCTA GACTCAAAAG GTCAGGTCTT GAAGTTGGCT 1080

GAAATCTTCA ATGCTCAAGA TATTTCCTTT AAGCAAATCC TTCAAGATGG CAAAGAGGGT 1140

GACAAGGCGC GTGTCGTTAT CATCACACAC AAGATTAATA AAGCCCAGCT TGAAAATGTC 1200

TCAGCTGAAT TGAAGAAGGT TTCAGAATTC GACCTCTTGA ATACCTTCAA GGTGCTAGGA 1260

GAATAAGATG AAGATTATTG TACCTGCAAC CAGTGCCAAT ATCGGGCCAG GTTTTGACTC 1320

GGTCGGTGTA GCTGTAACCA AGTATCTTCA AATTGAGGTC TGCGAAGAAC GAGATGAGTG 1380

GCTGATTGAA CACCAGATTG GCAAATGGAT TCCACATGAC GAGCGTAATC TCTTGCTCAA 1440

AATCGCTTTG CAAATTGTAC CAGACTTGCA ACCAAGACGC TTGAAAATGA CCAGTGATGT 1500

CCCTTTGGCG CGCGGTTTGG GTTCTTCCAG CTCGGTTATC GTTGCTGGGA TTGAACTAGC 1560

CAACCAACTG GGTCAACTCA ACTTATCAGA CCATGAAAAA TTGCAGTTAG CGACCAAGAT 1620

TGAAGGGCAT CCTGACAATG TGGCTCCAGC CATTTATGGT AATCTCGTTA TTGCAAGTTC 1680

TGTTGAAGGG CAAGTCTCTG CTATCGTAGC AGACTTTCCA GAGTGTGATT TTCTAGCTTA 1740

CATTCCAAAC TATGAATTAC GTACTCGCGA CAGCCGTAGT GTCTTGCCTA AAAAATTGTC 1800

TTATAAGGAA GCTGTTGCTG CAAGTTCTAT CGCCAATGTA GCGGTTGCTG CCTTGTTGGC 1860

AGGAGACATG GTGACCGCTG GGCAAGCAAT CGAGGGAGAC CTCTTCCATG AGCGCTATCG 1920

TCAGGACTTG GTAAGAGAAT TTGCGATGAT TAAGCAAGTG ACCAAAGAAA ATGGGGCCTA 1980

TGCAACCTAC CTTTCTGGTG CTGGGCCGAC AGTTATGGTT CTGGCTTCTC ATGACAAGAT 2040

GCCAACAATT AAGGCAGAAT TGGAAAAGCA ACCTTTCAAA GGAAAACTGC ATGACTTGAG 2100

AGTTGATACC CAAGGTGTCC GTGTAGAAGC AAAATAAAGA ATAGAAGATA GGATGGGGAA 2160

ACTCTTGACC AGAGGGGTTC ATATCCTTTT TGTGAAAAGA AGTTTATACT CAATGAAAAT 2220

CAAAGAGCAA ACTAGGAAGC TAGCCGCAGG CTGCTCAAAA CAGTGTTTTG AGGTTGCAGA 2280

TAGAACTGAC GAAGTCAGCT CAAGACACTG TTTTGAGGTT GCAGATAGAA CTGACGAAGT 2340

CAGTAACCAT ACTACGGTAA GGTGACGCTG ACGTGGTTTG AAGAGATTTT CGAAGAGTAT 2400

TAGTTAAAAA CGTGATAAAG GAGAAATAAA GATGGCAGAA ATTTATCTAG CAGGTGGTTG 2460

TTTTTGGGGC CTAGAGGAAT ATTTTTCACG CATTTCTGGA GTGCTAGAAA CCAGTGTTGG 2520

CTACGCTAAT GGTCAAGTCG AAACGACCAA TTACCAGTTG CTCAAGGAAA CAGACCATGC 2580

AGAAACGGTC CAAGTGATTT ACGATGAGAA GGAAGTGTCA CTCAGAGAGA TTTTACTTTA 2640

TTATTTCCGA GTTATCGATC CTCTATCTAT CAATCAACAA GGGAATGACC GTGGTCGCCA 2700

ATATCGAACT GGGATTTATT ATCAGGATGA AGCAGATTTG CCAGCTATCT ACACAGTGGT 2760

GCAGGAGCAG GAACGCATGC TGGGTCGAAA GATTGCAGTA GAAGTGGAGC AATTACGCCA 2820

CTACATTCTG GCTGAAGACT ACCACCAAGA CTATCTCAGG AAGAATCCTT CAGGTTACTG 2880

TCATATCGAT GTGACCGATG CTGATAAGCC ATTGATTGAT GCAGCAAACT ATGAAAAGCC 2940

TAGTCAAGAG GTGTTGAAGG CCAGTCTATC TGAAGAGTCT TATCGTGTCA CACAAGAAGC 3000

TGCTACAGAG GCTCCATTTA CCAATGCCTA TGACCAAACC TTTGAAGAGG GGATTTATGT 3060

AGATATTACG ACAGGTGAGC CACTCTTTTT TGCCAAGGAT AAGTTTGCTT CAGGTTGTGG 3120

TTGGCCAAGT TTTAGCCGTC CGATTTCCAA AGAGTTGATT CATTATTACA AGGATCTGAG 3180

CCATGGAATG GAGCGAATTG AAGTTCGTTC TCGTTCAGGC AGTGCTCACT TGGGTCATGT 3240

TTTCACAGAT GGACCGCGGG AGTTAGGCGG CCTCCGTTAC TGTATCAATT CTGCTTCTTT 3300

ACGCTTTGTG GCCAAGGATG AGATGGAAAA AGCAGGATAT GGCTATCTAT TGCCTTACTT 3360

AAACAAATAA AACAGAGAGT GGGGCTTCCC ACTTTCTTCA TTTCTAGAAT ATGAATAGAA 3420

GGGATTTATG AAACACCTAT TATCTTACTT CAAACCCTAC ATCAAGGAAT CAATTTTAGC 3480

CCCCTTGTTC AAGCTGTTAG AAGCTGTTTT TGAGCTCTTG GTTCCCATGG TGATTGCTGG 3540

GATTGTTGAC CAATCTTTAC CTCAGGGAGA TCAAGGTCAT CTCTGGATGC AGATTGGCCT 3600

GCTCCTTATC TTTGCAGTAA TTGGCGTTTT AGTGGCCTTG ATAGCTCAAT TTTACTCAGC 3660

AAAGGCAGCA GTAGGTTCTG CTAAGGAATT GACAAACGAT CTTTATCGTC ATATTCTTTC 3720

CTTGCCCAAG GACAGCAGAG ACCGTCTGAC AACTTCTAGT TTGGTCACTC GCTTGACTTC 3780

GGATACCTAC CAGATTCAGA CTGGTATCAA TCAATTCCTG CGTCTCTTTT TACGAGCGCC 3840

CATTATCGTT TTTGGTGCCA TTTTTATGGC TTATCGAATC TCAGCTGAGT TGACTTTCTG 3900

GTTCTTAGTC TTGGTTGCCA TTTTGACCAT TGTCATTGTA GGGTTATCTC GATTGGTCAA 3960

TCCTTTCTAC AGTAGTCTCA GAAAGAAAAC GGACCAACTG GTTCAGGAAA CGCGCCAGCA 4020

ATTGCAAGGG ATGCGGGTTA TTCGTGCTTT TGGTCAAGAA AAACGAGAGT TACAGATTTT 4080

TCAAACCCTT AACCAAGTTT ATGCTAGATT ACAAGAAAAG ACAGGTTTCT GGTCTAGTTT 4140

ATTAACACCT CTGACCTATC TGATTGTCAA TGGAACTCTT CTCGTTATTA TCTGGCAAGG 4200

CTATATTTCA ATTCAAGGAG GAGTGCTCAG TCAAGGTGCT CTCATTGCTC TTATCAATTA 4260

CCTCTTACAG ATTTTGGTGG AATTGGTCAA GCTAGCCATG TTGATCAATT CCCTCAACCA 4320

GTCCTATATC TCAGTCAAGC GAATCGAGGA AGTCTTTGTT GAGGCTCCAG AGGATATCCA 4380

TTCAGAGTTA GAACAAAAGC AAGCTACCAG AGATAAGGTT TTACAAGTCC AAGAATTGAC 4440

CTTTACCTAT CCTGATGCGG CCCAGCCTTC TCTGAGATAC ATTTCCTTTG ATATGACTCA 4500

AGGACAAATT CTAGGTATCA TCGGGGGAAC TGGTTCTGGT AAATCAAGCT TGGTGCAACT 4560

CTTACTTGGA CTTTATCCAG TAGACAAGGG GAACATTGAC CTTTATCAAA ATGGACGTAG 4620

TCCTCTTAAT TTGGAGCAGT GGCGGTCTTG GATTGCCTAT GTACCTCAAA AGGTCGAACT 4680

CTTTAAAGGA ACCATTCGTT CCAACTTGAC TCTAGGTTTC AATCAAGAAG TATCTGACCA 4740

GGAACTCTGG CAGGCCTTGG AGATTGCGCA AGCTAAGGAT TTTGTCAGTG AAAAGGAAGG 4800

ACTCTTGGAT GCTCTAGTTG AGGCAGGGGG GCGAAATTTC TCAGGTGGAC AAAAACAAAG 4860

ATTGTCTATC GCCCGAGCAG TCTTGCGCCA GGCTCCGTTT CTCATCCTAG ATGATGCAAC 4920

CTCGGCACTG GATACCATTA CAGAGTCCAA GCTCTTGAAA GCTATTAGAG AAAATTTTCC 4980

AAACACGAGC TTAATTTTGA TCTCTCAACG AACCTCAACT TTACAGATGG CGGACCAGAT 5040

TCTCCTCTTG GAAAAAGGTG AGTTGCTAGC TGTTGGCAAG CACGATGACT TGATGAAATC 5100

CAGCCAAGTC TATTGTGAAA TCAATGCATC CCAACATGGA AAGGAGGACT AGAATGAAAC 5160

GACAAACTGT AAACCAGACG CTCAAACGTT TAGCCGTAGA TTTAGCAAGC CATCCTTTCC 5220

TCCTTTTCCT AGCCTTTCTA GGAACTATTG CCCAAGTTGG CTTATCAATT TACCTACCTA 5280

TTCTGATTGG GCAGGTCATT GACCAAGTCC TAGTGGCTGG TTCATCACCA GTTTTTTGGC 5340

AGATTTTTCT CCAGATGCTC TTGGTGGTAA TAGGAAATAC TCTGGTACAA TGGGCCAATC 5400

CTCTCCTCTA TAATCGTCTA ATCTTCTCTT ATACCAGAGA TTTACGGGAG CGAATCATCC 5460

ATAAGCTCCA TCGTTTACCG ATTGCCTTTG TAGATAGGCA AGGTAGTGGA GAGATGGTTA 5520

GTCGTGTAAC CACGGACATC GAACAGTTGG CAGCTGGCTT GACCATGATT TTTAACCAAT 5580

TTTTCATTGG TGTTTTGATG ATTTTGGTCA GTATTCTAGC CATGCTCCAA ATTCATCTCC 5640

TCATGACTCT CTTAGTCTTG CTGTTGACGC CACTGTCCAT GGTGATTTCA CGCTTTATTG 5700

CCAAGAAATC CTATCATCTC TTCCAGAAGC AAACAGAGAC GAGGGGAATT CAGACTCAGT 5760

TGATTGAAGA ATCGCTTAGT CAGCAGACTA TAATCCAGTC CTTCAATGCT CAAACAGAAT 5820

TTATCCAAAG ATTGCGTGAG GCTCATGACA ACTACTCAGG CTATTCTCAG TCAGCCATCT 5880

TTTATTCTTC AACGGTCAAT CCTTCGACTC GCTTTGTAAA TGCACTCATT TATGCCCTTT 5940

TAGCTGGAGT AGGAGCTTAT CGTATCATGA TGGGTTCAGC CTTGACCGTC GGTCGTTTAG 6000

TGACTTTTTT GAACTATGTT CAGCAATACA CCAAGCCCTT TAACGATATT TCTTCAGTGC 6060

TAGCTGAGTT GCAAAGTGCT CTGGCTTGCG TAGAGCGTAT CTATGGAGTC TTAGATAGCC 6120

CTGAAGTGGC TGAAACAGGT AAGGAAGTCT TGACGACCAG TGACCAAGTT AAGGGAGCTA 6180

TTTCCTTTAA ACATGTCTCT TTTGGCTACC ATCCTGAAAA AATTTTGATT AAGGACTTGT 6240

CTATCGATAT TCCAGCTGGT AGTAAGGTAG CCATCGTTGG TCCGACAGGT GCTGGAAAAT 6300

CAACTCTTAT CAATCTCCTT ATGCGTTTTT ATCCCATTAG CTCGGGAGAT ATCTTGCTGG 6360

ATGGGCAATC CATTTATGAT TATACACGAG TATCATTGAG ACAGCAGTTT GGTATGGTGC 6420

TTCAAGAAAC CTGGCTCACA CAAGGGACCA TTCATGATAA TATTGCCTTT GGCAATCCTG 6480

AAGCCAGTCG AGAGCAAGTA ATTGCTGCTG CCAAAGCAGC TAATGCAGAC TTTTTCATCC 6540

AACAGTTGCC ACAGGGATAC GATACCAAGT TGGAAAATGC TGGAGAATCT CTCTCTGTCG 6600

GCCAAGCTCA GCTCTTGACC ATAGCCCGAG TCTTTCTGGC TATTCCAAAG ATTCTTATCT 6660

TAGACGAGGC AACTTCTTCC ATTGATACAC GGACAGAAGT GCTGGTACAG GATGCCTTTG 6720

CAAAACTCAT GAAGGGCCGC ACAAGTTTCA TCATTGCTCA CCGTTTGTCA ACCATTCAGG 6780

ATGCGGATTT AATTCTTGTC TTAGTAGATG GTGATATTGT TGAATATGGT AACCATCAAG 6840

AACTCATGGA TAGAAAGGGT AAGTATTACC AAATGCAAAA AGCTGCGGCT TTTAGTTCTG 6900

AATAAGCCAT TCTCTTTTGA AAGTTTATGG ACGAAAAAAG TTGCCTTCGA GTGACTTTTT 6960

TGTTACAATA GCTAGAAAAA TTGTTCACTG TAATACTCAA TGAAAATCAA AGAGCAAACT 7020

AGGAAGCTAG CCGTAGGTTG CTCAAAGCAC AGCTTTGAGG TTGTAGATAA GACTGACGAA 7080

GTCAGTTCAA AACACTGTTT TGAGGTTGCA GATAGAACTG ACGAAGTCAG CTCAAAACAC 7140

TGTTTTGAGG TTGCAGATAG AACTGACGAA GTCAGCTCAA AACAGG 7186

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14273 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

CTGAAAATTC TAAAAAATTT ATAAGTAAGG AATTAATTAG TTATTTTTGT GATAAAGTTT 60

390

ATGATGAAAT ATTTGTTGAA GAGGTAGTTC CGCACGTTTT TCTGCCATAT GAATCTGACT 120

TACTTCTTAT TTTACCAGCT ACGGCAAATG TGATTGGCAA AATTGCTAAT GGTATTGCTG 180

ATGATTTAGT TACAGCAACT GTTTTAAACT TTAATAAAAA AATAATTTTT TGTCCCAATA 240

TGAACTCTAC TATGTGGGAC AATCACATAG TTCAAAGAAA TGTATCAATT CTAAAGGAGT 300

TGGGACATAT ATTTTTATTT GAGTCTAAAA AAACATATGA GGTAGGATTG CGTAAAGCAA 360

TAGATTCAAC ATGTTCAATG TTACAACCAC AGTCGTTAGT AAAAGAACTT ATCAAATTAG 420

AAAATATTGT CCTTGAAGAG GGACATTAAA AACTACTGAG AATATTAATG AGGGGAAAAA 480

ATGGAAAATT CATCAATCGA TGTAGATATG CTGTTGGAAG AATTGACACA AGAAGCAATG 540

GTCGTTGTTG CTGTTGATAA GGACTGTTAA TTTAAACTTA TGGCAATATA TGAAAGGTTA 600

CTGGATGTTT TAAATTATGC AGGCAGTAGC CTTTTATTAT ATACAAATGG ATAAAGTAAG 660

GATAATACAA TGATTAATAA AAAAATACAA CAAGTTGTTT TGGAATCATT ACAGAATTTT 720

TTGAATGGGA ACTTCATTTC GCCTTGTGTA GTCTATGATT TTGGCTTGCT GGAAACTGTA 780

CTTGATGAAT TTAAAAATCA AATTCCTGTA ACATTCAATT ACCAACTTTT TTATGCCGTT 840

AAAGCAAATT CAAATGAGAA GATACTTGAA TTCTTAGTAG ATAAAATTGA TGGAGTTGAT 900

GTGGCGTCAT TATCTGAATT AGATGTGGCT AAAAAATTTT TCCCACCAAC TCAAATTTCT 960

GTTAATGGTC CCGCATTTTC TTATGAAACT TTATATAATC TGATTAAAAA ACAATATAAA 1020

GTTGATATTA ACTTTTTGGA ACATCTTCAA CAATTTTCCC CAAAAGAATC TGTTGGAATA 1080

AGAGTAACGG AGCCAGATGA ACTTAATAAT CGTATGAGTC GATTTGGAAT AAATATTTGC 1140

AGTGATAATT GGACTAGTAA TTTACAAAAT CCTTTAATTA CACGACTGCA TTTTCATTTT 1200

GGAGAAAAAG ATGATAAATT TATTGTTAAG TTAGATAAAA TATTATTTAA GTTACAAGAA 1260

ATTAATAAAC TTAGAGAGGT TAGAGAAATA AATCTTGGAG GCGGTTTTAT GAAATTATTT 1320

ATGGAAAATC GTTTGAAAGA ATTTTTTCTA TCACTTATGG AAATCTATAA AAAGTACGAT 1380

ATTGATAGTA CTGTGACTAC AATAATAGAA CCAGGTAGTG CAATTACTTC ATTTTCTGCC 1440

TATATGATTA CTAGCCCAGT TAATGTTAGT GAGGTGAATG AGCAGCAGGT TATCACGTTA 1500

GACACATCAA TATACACCAA TACATTATGG TTTGTTCCGC ATATTATTAC AACGTTAAAT 1560

TCAAGTAGTA AAGAGCGTTA TAGTACTATT CTCTATGGTA ATACCTGTTA TGAACATGAC 1620

AAGTATAAAA TGAAAGTTTC GCTTCCAAGG TTAACTCAAA ATAGCAGTAT AGTGTTTTTT 1680

CCTGTAGGAG CTTATATAAA AAGCAATCAT TCAAATTTAC ATCGTAATGA TTTTATGCGG 1740

GAGGTATATT TGTGGACAAA AAACTTGACA TATTAGATAA AGTTAAGGAA TATTTAGGAA 1800

ATAAAACTAC TCAAATTCTG GATAATCAAT ATAAAGAATT TTTGAAACTT AATGATATAA 1860

GGCGAGCGTT TGGTATTTCA GAAAAAGTAT TAAACAATTC TTTTAATTTT ACGAGTAAAG 1920

AATTTAATGA TTTAATTAAT AACGAAAATT ATTTATTCGA ATATGCATGT AGAATTAGAG 1980

AGGAATGGAG AAAAAAATGC TTTAATCATT CTTATCGTTT TCTATGCTCA CCTATAATTA 2040

CAGATGATTT TCTTAACACG AAGACATTGA GAAGTAGCCA AATTGAATAT AAATATGAGC 2100

GATATTTATC GAAAAGTTCG ATAGGCGATA GAGCGGTTGA TGGCTTTGTT TCCTTCAATA 2160

CTTTAACAGC TAATGGTATG TCTGCTATTA AACTATGTCT TGAGATATTA AACTCTATTT 2220

TCTTCAAGAA GAAGATTGAT TTATTATATT CAACCGGATA TTATGAAACA AGATTTTTAT 2280

TAAATAATCT TGCTAAATCA GGTATTAGTT GCTATGAGGT AAGTAATTGT GAATTGGATA 2340

AAGATAAATT TTATAATGTA TTCATGATGG AACCCAATCG AGCCGATTTA ACATTACAAA 2400

AAACTGATTT CAAGATAGTA GAATATTTTG TTAAGTATAA AAATAATTCA ATAAAAGTCG 2460

TTATTTTAGA TATTTCATAT CAAGGTTCTA ATTTTAAATT AGTAGAATTT TTAGAGAAAT 2520

TTAAATTTGC GAATGTAATT ATTTTTGTGG TACGATCTTT GATAAAATTA GATCAAATGG 2580

GATTAGAATT GACAAATGGG GGAATAATAG AAGTGTTTAT TCCTAATCAT TTGAGAAAGT 2640

TGAAAAATTT TATTGAAGAG GAATTCAATA AATTTAGAAA TTCTCACGGA GCTAATCTAA 2700

GCCTCTATGA ATACTGTTTG CTTGATAATT CTTTAACTTT AAAAAATGAT TGGAACTATT 2760

CTGATTTAGT TATGAAATTT ACGAGTAATT TTTATGCTGA TATAAAAGAC TTGTTCATGG 2820

AAAATTCTGA TATTGAAATC ATCCATGAAG AGGGAGTACC TTTTGTATTT TTAGATTTAA 2880

TAGGTGAAGG TAAAAAAAGAA TATGAAATGT TTTTTCAATG GTTAAACTTC TTTTACAAAC 2940

AGCTTGGAAT CACATTGTAT GCTAGAAATA GTTTTGGGTT TCGGAATCTA ACAGTAGAGT 3000

ATTTTGGAAT TATTGGGACA GAAAGATATA TATTTAAGAT TTGTCCAGGT GTTTATAAAG 3060

GGTTAAGTTA TTATTTGATG AAATTTTTAT TAAAATCTTT TTCAAATGAA TATTTAAAAA 3120

CTACTGATGA GGTTAATAGA TGAAAAATTT GATAAAGTTG CTAATAATTA GATTGATTGT 3180

TAACTTAGCA GACAGTGTAT TTTATATAGT AGCATTGTGG CACGTTAGCA ATAATTATTC 3240

TTCGAGCATG TTCTTAGGAA TATTTATTGC AGTAAATTAT CTACCGGATT TGTTACTAAT 3300

CTTTTTTGGA CCAGTTATTG ACAGAGTAAA TCCGCAAAAA ATTCTTATAA TATCAATTTT 3360

GGTTCAATTA GCAGTGGCTG TAATATTTTT ATTATTATTA AACCAAATAT CATTTTGGGT 3420

GATAATGAGT CTAGTGTTTA TTTCAGTAAT GGCTAGCTCC ATAAGTTACG TGATAGAAGA 3480

TGTGTTGATT CCTCAAGTGG TAGAATATGA TAAGATTGTA TTTGCAAATT CTCTTTTTAG 3540

TATTTCGTAT AAAGTATTAG ATTCTATTTT TAATTCATTC GCATCATTTT TACAGGTGGC 3600

AGTAGGATTT ATTTTATTGG TTAAGATAGA TATAGGCATA TTTTCACTTG CTCTATTTAT 3660

ATTGTTGTTG TTAAAATTTA GAACTAGCAA TGCGAATATA GAAAACTTCT CTTTCAAATA 3720

TTACAAGAGA GAAGTGTTGC AAGGTACAAA GTTTATTTTA AATAATAAAT TATTATTTAA 3780

AACCAGTATT TCTTTAACGC TTATAAACTT TTTTTATTCA TTTCAGACAG TAGTTGTACC 3840

GATTTTTTCT ATTCGATATT TTGATGGTCC GATTTTTTAT GGTATTTTTT TAACTATTGC 3900

TGGTTTGGGT GGTATATTGG GAAATATGCT AGCGCCAATC GTAATAAAAT ATTTAAAATC 3960

GAATCAAATT GTTGGTGTAT TTCTTTTTTT GAACGGCTCA AGTTGGTTAG TAGCAATTGT 4020

TATAAAAGAC TATACTTTAT CACTTATTTT ATTTTTCGTT TGTTTTATGT CTAAAGGAGT 4080

CTTCAATATT ATTTTTAATT CGTTGTACCA ACAAATACCT CCACATCAAC TTCTTGGTAG 4140

GGTAAATACT ACCATTGATT CTATTATTTC TTTTGGAATG CCAATTGGTA GTTTAGTTGC 4200

AGGAACGCTT ATTGATTTGA ATATTGAATT AGTGTTAATT GCTATTAGCA TACCTTATTT 4260

TTTGTTTTCT TATATTTTTT ATACGGATAA TGGATTGAAA GAATTTAGTA TATATTAGAA 4320

ATGTTTATGT TCATTCAAAA GCATAATGAC TATAACTGAA AAAGAAAAGT GATATCTTTA 4380

AGGTTGTTCT TCTTGGTGGT GAGATTCGTG AGACAACCCA AGCTTTTGTC GGAAAGATTA 4440

CCAATGCTTT GATGGATAGG ATGTACTTTA GCAAGATGTT TTTAGTGGTA ACGGTATCGT 4500

GGATGGACGT GTAATAACCT CTTCTTTCGA GGAGTATTTT ACTAAAAAAC TAGCCTTGGA 4560

GCGTTCCCCA GAAACGGACT TACTCATTGA CTCTTCAAAG ATTTGGGGAG AAGATTTTGC 4620

TTCATCTGTT CCTTGAAAAA AGTCACAGCA GTCATCACAG ACGATAGTAC TGAACAAAAC 4680

TATGAAGAGT TAGAAATTTA TACGCAGGTG ATTGTATAAA GGATCTGGAA ATAGATAAGA 4740

AGTTGATTAG TATTGACCTA GGTGGTACAA ATATTAAGAT TACTGTTCTT TCAAATGACG 4800

GTGAGATTGA AACTTTGTGG AGTATTACAA CAGATACAAG TGAGAAAGGT TCTCAAATTA 4860

TATCGGACAT CATCAGTTCT ATTAAAAATA AATTGACCGA ACGGAATATT CCTGATAGCG 4920

ACCTTCTTGG AATCGGTATG GGAAGTTGCT CATCATACTT TCCTTGTAAA TCATAGGGGC 4980

TATAAACTCT CCGTCTACTT GTCCTGCAAC AATTGAAGTC TGCTCAAAAC GCCGTCCGCT 5040

AATCTTTTCA TAGACTTTCT CCCTTTTAGG AGCCTAGCTT TCTAGTTTGT TCTTTGATTT 5100

TTATTGAGTA TACCACTATT TTACTCCCTC TGGCAAGGGA CTTTGTCTAT GTGGAGGGAT 5160

TGGGCTCCTA TGTGGTGGAG CTTTTCTGTT CTTTCTGAAA TATGGTATAA TAGCACTAAT 5220

CAATTTCTAG GAAAATAGAT ACAGAAAGGG GCTGAAAGAT GTCTCATATT ATTGAATTGC 5280

CAGAGATGCT GGCAAACCAA ATCGCGGCTG GAGAGGTCAT TGAACGTCCT GCCAGTGTGG 5340

TCAAAGAGTT GGTAGAAAAT GCCATTGACG CGGGCTCTAG TCAGATTATC ATTGAGATTG 5400

AGGAAGCTGG TCTCAAGAAG GTTCAAATCA CGGATAACGG TCATGGAATT GCCCACGATG 5460

AGGTGGAGTT GGCCCTGCGT CGCCATGCGA CCAGTAAGAT AAAAAATCAA GCAGATCTCT 5520

TTCGGATTCG GACGCTTGGT TTTCGTGGTG AAGCCTTGCC TTCTATTGCG TCTGTTAGTG 5580

TCTTGACTCT GTTAACGGCG GTGGATGGTG CTAGTCATGG AACCAAGTTA GTCGCGCGTG 5640

GGGGTGAAGT TGAGGAAGTC ATCCCAGCGA CTAGTCCTGT GGGAACCAAG GTTTGTGTGG 5700

AGGATCTCTT TTTCAACACG CCTGCCCGTC TCAAGTATAT GAAGAGCCAG CAAGCGGAGT 5760

TGTCTCATAT CATTGATATT GTCAACCGTC TGGGCTTGGC CCATCCTGAG ATTTCTTTTA 5820

GCTTGATTAG TGATGGCAAG GAAATGACGC GGACAGCAGG GACTGGTCAA TTGCGCCAAG 5880

CAATCGCAGG GATTTACGGT TTGGTCAGTG CCAAGAAGAT GATTGAAATT GAGAACTCTG 5940

ACCTAGATTT CGAAATTTCA GGTTTTTGTGT CCTTGCCTGA GTTGACTCGG GCTAACCGCA 6000

ATTATATCAG CCTCTTCATC AATGGCCGTT ATATTAAGAA CTTCCTGCTC AATCGTGCTA 6060

TTTTGGATGG TTTTGGAAGC AAGCTTATGG TTGGACGTTT TCCACTGGCT GTCATTCACA 6120

TCCATATCGA CCCTTATCTA GCGGATGTCA ATGTGCATCC AACTAAGCAA GAGGTGCGGA 6180

TTTCCAAGGA AAAAGAACTG ATGACTCTGG TTTCAGAAGC TATTGCAAAT AGTCTCAAGG 6240

AACAAACCTT GATTCCAGAT GCCTTGGAAA ATCTTGCCAA ATCGACCGTG CGCAATCGTG 6300

AGAAGGTGGA GCAAACTATT CTCCCACTCA AAGAAAATAC GCTCTACTAT GAGAAAACTG 6360

AGCCGTCAAG ACCTAGTCAA ACTGAAGTAG CTGATTATCA GGTAGAATTG ACTGATGAAG 6420

GGCAGGATTT GACCCTGTTT GCCAAGGAAA CCTTGGACCG ATTGACCAAG CCAGCAAAAC 6480

TGCATTTTGC AGAGAGAAAG CCTGCTAACT ACGACCAGCT AGACCATCCA GAGTTAGATC 6540

TTGCTAGCAT CGATAAGGCT TATGACAAAC TGGAGCGAGA AGAAGCATCC AGCTTCCCAG 6600

AGTTGGAGTT TTTCGGACAA ATGCACGGGA CTTATCTCTT TGCCCAAGGG CGAGATGGAC 6660

TTTACATCAT AGATCAGCAC GCTGCTCAGG AACGGGTCAA GTACGAGGAG TACCGTGAAA 6720

GCATTGGCAA TGTTGACCAA AGCCAGCAGC AACTCCTAGT GCCCTATATC TTTGAATTTC 6780

CTGCGGATGA TGCCCTGCGT CTCAAGGAAA GAATGCCTCT CTTAGAGGAA GTGGGCGTCT 6840

TTCTAGCAGA GTACGGAGAA AATCAATTTA TTCTACGTGA ACATCCTATT TGGATGGCAG 6900

AAGAAGAGAT TGAATCAGGC ATCTATGAGA TGTGCGACAT GCTCCTTTTG ACCAAGGAAG 6960

TTTCTATCAA GAAATACCGA GCAGAGCTGG CTATCATGAT GTCTTGCAAG CGATCTATCA 7020

AGGCCAATCA TCGTATTGAT GATCATTCAG CTAGACAACT CCTCTATCAG CTTTCTCAAT 7080

GTGACAATCC CTATAACTGT CCTCACGGAC GTCCTGTTTT GGTGCATTTT ACCAAGTCGG 7140

ATATGGAAAA GATGTTCCGA CGTATTCAGG AAAATCACAC CAGTCTCCGT GAGTTGGGGA 7200

AATATTAAAA GTATAAAAAA GTCTGGGAAA AATTTTCAAA ATCAAAAAAA CGCATAAAAT 7260

CAGGTGTTCA AAAACCTTGA TTTTATGCGT TTTATCATGG AAATAGTTAC TTCATTTTTT 7320

CCTAATTCTT TTCGAAACTC TTTTTAAACG ACGTCAGTTT TATCAGTAAT CTCAAAACAG 7380

TGTTTTGAGC TAATTTTGCC AGTTTTGTCT GTAACATCGA AGTTGTGTTT TACCACTCTG 7440

CGACTGGTTT CCTAGTTTGC TCTATGATTT TCACAGAGCA TTAAATTGCG ATTTTGCCAA 7500

GTTTCTTTAT TCGTCTAAAA GTAGAGTCTG TTCTATGCGT CTAATGTACG AATCAGGTTG 7560

ACCATTTCAA TAGCTCCTTG TGCACACTCA GAACCCTTAT TTCCTGCTTT AGTACCAGCT 7620

CGTTCTATGG CTTGTTCAAT TGTATCTGTC GTTAGCACAC CAAACATAAC AGGAATTTCG 7680

CTATTTAAAC TGATTTGGGC GATTCCCTTA GATACCTCGC TACATACATA ATCATAATGA 7740

CTTGTATTCC CTCTAATGAC AGCTCCCAAG CAGATAATTG CATCATATTT TTTACTTTTT 7800

GCCATTTTTG ATGCAATCAG TGGTATTTCA AAAGCTCCTG GAACCCAGGC TACCTCTATA 7860

TCTTTCTCGT TTACATTCTC TCTTTTGAGA TTATCTAGTG CTCCAGATAA TAATTTTGAA 7920

GTTATAAATT CATTAAATCT CGCTACAACA ATACCTATTT TAATATTGTT TGCTACTAAA 7980

TTACCTTCAT AAGTGTTCAT TTATTTTTCC TCCATATTTA AAATGTGACC CATTCGATTT 8040

TTCTTTGTTT CTAAATAAAA ACTATCGTAA GGATTGGCTT CTATTTCGAT TGATATTCTA 8100

CTGGAAATGG TAATTCCATA TTTTTCTAAC TGTTCAACCT TGTCAGGATT ATTTGTCAGT 8160

AAATGAAGTG ACTGAAGTCC CAGATCTTTA AGCATTTTTG CTCCAATATG ATATTCTCTT 8220

AAATCACCTT CAAAGCCTAA TGCAAGATTG GCATCAAGCG TATCCATGCC TTGATCTTGT 8280

AAATGATAGG CTTTTAATTT ATTGATAAGT CCAATTCCTC GTCCCTCCTG TCGCAAGTAA 8340

AGTAAGACAC CCGAACCATT CTCAACAATC ATTTTCATAG CCTTATCGAA TTGCTGTCCA 8400

CAATCGCAAC GTAAAGAGCC TAAAACATCT CCTGTTAAAC ATTCGGAGTG GACCCGACAT 8460

AATACATTGG CTTCATCCTC TATATTTCCC ATAATAAGAG CAAGATGATG TTCCCCATTT 8520

AGTTTATCTA TATAGCTAAT TGCTTTGAAA TTACCGTATC TAGTAGGCAT ATTGACAGTT 8580

GAAACTCGTT CTACCAGCTG ATCATATACT TTTCTATATT CTTGTAATTC TTTGATGGTA 8640

ATTAGTGGAA TGTTGTGTTT TTTCGAGAAC TGAATTAAAT CATCTGTTCT CATCATTTTG 8700

CCATCATGAT TCATTATTTC ACAACATAGG CCACACTCTT TTAGTCCAGC TAATTTTAAT 8760

AAATCAACAG TTGCTTCTGT GTGTCCATTT CTTTCTAGGA CACCACCTTT TTTTGCAATT 8820

AAAGGAAACA TGTGTCCTGG CCTGCGAAAA TCAGAGGGTG TTATATCTTC AGCTACACAC 8880

ATACGTGCGG TCAGTCCTCT TTCCTCGGCA GAAATACCTG TGGTCGTTTC TTTATAATCA 8940

ATTGAAACTG TAAAAGCAGT CTTATGATTA TCTGTATTGT TTTCAACCAT AGGTGAAAGC 9000

ATTAATTGAT TAGCTAAACT TTCGCTCATA GGCATACAAA TTAATCCTTT GGCATAAGTA 9060

GCCATAAAAT TAACATTTTC TGTTGTAGCT GCTTGTGCAG AACAAATTAA GTCTCCTTCA 9120

TTTTCTCTAT CCTTGTCGTC TATAACAAGA ACAAGTCGTC CCTTCTGCAA TGCTTCTAAT 9180

GCTTCTTGTA TTTTTCGATA TTCCATTGAC TGATTATCCT TTCTGCTAAA ATCCATTTTG 9240

ATATAATAGT TCCTTAGATA TTTCTGATTT TGGAGAGTTA TCCATCAGTT TTTGCACATA 9300

TTTACCTAAG ATATCATTTT CAAGATTTAC TGTACTCCCG ACTTGTTTAC TCTTAAGAAT 9360

GGTTTGTTCC AAGGTATGAG GGATAACAGA TACTGAAAAG TTTACTTTGG AGACTTTAGC 9420

GACAGTCAGA CTAATGCCGT CAATTGTAAT AGATCCTTTT TCAACTATTA AATCTAAAAT 9480

TTCTTTTTGT GTGTTGATTT GATACCATAC AGCATTATCA TCTTTTTTTA TTGACGAGAT 9540

TTTTCCTGTA CCATCAATGT GTCCTGTAAC GACGTGACCC CCAAGTCGAC CGTTGACAGA 9600

TAAGGCTCTT TCTAGATTCA CCTCACTTCC ATGTTTTAAT AGAGTAAGAG CTGTTCGACT 9660

CCATGTTTCA TTCATTACAT CAACTGTAAA GGATTGATGA TTGAAATGAG TAACTGTAAG 9720

ACAGATACCA TTTACTGCTA TACTATCGCC TAAATGGATA TCCGTTAATA TTTTTGAGGC 9780

TTTAATTGAT AGTTTACAAT TACGAGAGTC TTTCTGTATT CTTTCAACTT TTCCGATTTC 9840

TTCAATTATT CCTGTGAACA TGGATAAATC ACTTCACTTT CTATGAGATA GTCATTTCCT 9900

ATTTGAGAAA ATGCATAAGG TTTCAATCTA ATAGCGTCAT TTGGCAAAGA AATACCTTCA 9960

CCTCCGACAG GAAACTTGGC ACTACCTCCA AAAACTTTTG GTGCAATATA TATTTTCAGC 10020

TCATCAACAA TTTGTTGTTC CAAAGCACTC CAATTCATTA GACTGCCCCC TTCTAGAACT 10080

AGGCTATCAA TCTGCATGTT TCCTAGATGT TGCATTAAAC TCGATAAGTC TATATGATTG 10140

CCTTTTTTCT TTATGGAAAG TATTTCACAG CCATGATTTT GATATAGCTT CATTTTATTT 10200

TTGTCTTCAG AGGAAGTGGC AATGTAAGTT TTAATATCAT TTGCTGTTTT TACGATTTTA 10260

GAGGTAAGAG GAGTTCGTAA ATGTGTATCG CATATGATAC GGATAGGATT TTTCCCTTCC 10320

TCCAATCTAC ATGTCAGCAA AGGATCGTCT TGAATAACAG TATTGACTCC CACCATAATT 10380

GCACTAACAT GGTGTCGTAA CTGATGCACA TGCTTTCTTG CTTCTTCTTC AGTAATCCAT 10440

TTGGATTGAT TTGTTTTAGT GGCTATTTTT CCATCCATTG ACATTGCATA TTTCATAAAA 10500

ACATAGGGTA CATGCTGGGT AATATACTTT CTAAAACTTT TTATTAAGTT AAGACACTCA 10560

TTTTCTAAAA TTCCAACAGT AACTTGAAGA TTATTTTCCT CAAGTATCTT TACTCCTTTT 10620

CCAGATACAA TAGGATTACA GTCTAGGCTT CCAATGACTA CTCTTGTAAT ACCACTATCG 10680

ATTATAGCAT CTATACAGGG AGGTGTTTTC CCGAAGTGAC AACAGGGTTC AAGTGTTACA 10740

TAAAGCGTCG CTCCGACAGG GGATTCTCTA CAGTTTTTAA GAGCATTTCT CTCAGCATGT 10800

GGGCCACCAA AAAACTCATG ATAACCTTGT CCGATAATGT GATTATCTTT TACAATAACT 10860

GCGCCGACCA TAGGATTGGG ATTGACGTAA CCAGCCCCTT TTTGTGCCAG TTTTATTGCT 10920

AATTTCATAT ATTTTGAATC GCTCATCTCG CTACCTCCAA AAAAATATAC CTTGAATAGG 10980

GGACTACTCA AGGCATACAA AAGAAAACTT ATGCGATTAA CAAAAATGCT CTGAAATGAC 11040

AAGTAATCAT TTCAGAGCAC GCAAAAAGCA CAAATATACT TTTATCTTCT TTCATCCAGA 11100

CTATACTGTC GGCTTTGGAA TTTCACCAAA TCATGCCTTT CGGCTCGTGG GCTATACCAC 11160

CGGTAGGGAA TTTCACCCTG CCCTGAAGAT AGTTATTCAA TTACAGATGA TTATAGTACT 11220

TAATTTTGAA TATGTCAACA GATAAATACC GATTGTTTTT GATATACTGT ATTTGTGATA 11280

ATCGATTCTC GCTCCTCGGA TAAAGAAAAT ATGATATACT AGATAAACGA AATAAGAGAG 11340

AAGGAATACT ATGTACGCAT ATTTAAAAGG AATCATTACC AAAATTACTG CCAAATACAT 11400

TGTTCTTGAA ACCAATGGTA TTGGTTATAT CCTGCATGTG GCCAATCCTT ATGCCTATTC 11460

AGGTCAGGTT AATCAGGAGG CTCAGATTTA TGTGCATCAG GTTGTGCGTG AGGACGCCCA 11520

TTTGCTTTAT GGATTTCGCT CAGAGGATGA GAAAAAGCTC TTTCTTAGTC TGATTTCGGT 11580

CTCTGGGATT GGTCCTGTAT CAGCTCTTGC TATTATCGCT GCTGATGACA ATGCTGGCTT 11640

GGTTCAAGCC ATTGAAACCA AGAACATCAC CTACTTGACC AAGTTCCCTA AAATTGGCAA 11700

GAAAACAGCC CAGCAGATGG TGCTGGACTT GGAAGGCAAG GTAGTAGTTG CAGGAGATGA 11760

CCTTCCTGCC AAGGTCGCAG TGCAAGCAAG TGCTGAAAAC CAAGAATTGG AAGAAGCTAT 11820

GGAAGCCATG TTGGCTCTGG GCTACAAGGC AACAGAGCTC AAGAAAATCA AGAAATTCTT 11880

TGAAGGAACG ACAGATACAG CTGAGAACTA TATCAAGTCG GCCCTTAAAA TGTTGGTCAA 11940

ATAGGAGCAG AGAATGACAA AACGTTGTTC GTGGGTCAAG ATGACCAACC CGCTCTACAT 12000

CGCCTATCAT GATGAGGAGT GGGGCCAGCC CCTCCATGAT GACCAAGTAT TGTTTGAGTT 12060

GTTGTGTATG GAAACCTATC AGGCAGGCCT GTCTTGGGAA ACGGTACTCA ACAAACGCCA 12120

AGCTTTCCGA GAAGTCTTTC ATAGCTATCA AATTCACTCA GTCGCAGAGA TGACTGACAC 12180

TGAATTGGAA GCCATGCTGG AGAATCCAGC TATCATTCGA AATAGAGCCA AGCTTTTTGC 12240

TACACGCGCT AACGCCCAAG CCTTTCTACA GTTACAGGCA GAGTACGGCT CTTTTGATGC 12300

CTATCTTTGG TCTTTTGTTG AGGGGAAAAC TGTCGTTAAC GATGTTCCTG ATTATCGCCA 12360

AGCGCCAGCT AAAACACCCT TATCTGAGAA ATTAGCCAAA GATCTCAAAA AACGAGGCTT 12420

CAAGTTCACA GGCCCAGTCG CCGTATTGTC TTTTCTACAG GCTGCAGGGC TAGTTGATGA 12480

CCACGAGAAT GATTGTGAGT GGAAAGGTCT TAAATGATGT CTAACAAAAA TAAGGAAATT 12540

CTGATTTTTG CGATTCTCTA TACAGTCCTC TTTATGTTTG ATGGCGTTAA ATTGCTGGCT 12600

TCTTTAATGC CATCTGCCAT TGCAAATTAT CTTGTTTATG TAGTTTTAGC TCTATATGGC 12660

TCCTTCTTGT TCAAGGATAG ATTGATCCAA CAATGGAAGG AGATTAGAAA GACTAAAAGA 12720

AAATTCTTCT TTGGAGTCTT AACAGGATGG CTCTTTCTCA TTCTGATGAC TGTTGTCTTT 12780

GAATTTGTAT CAGAGATGTT GAAGCAGTTT GTGGGACTAG ATGGACAAGG TCTAAATCAG 12840

TCTAATATTC AAAGTACCTT TCAAGAACAA CCACTACTGA TAGCTGTTTT TGCTTGTGTC 12900

ATTGGACCTC TGGTAGAAGA ATTATTTTTC CGTCAGGTCT TATTGCATTA CTTGCAGGAA 12960

CGGTTGTCAG GTTTACTAAG CATTATTCTG GTAGGACTTG TTTTTGCTCT GACTCATATG 13020

CACAGTTTGG CTCTATCAGA GTGGATTGGT GCAGTTGGTT ACTTAGGTGG AGGCCTTGCC 13080

TTTTCTATTA TTTATGTGAA AGAAAAAGAG AATATCTACT ATCCCCTACT TGTTCACATG 13140

TTAAGCAACA GCCTCTCCTT AATCATTTTA GCTATCAGTA TAGTAAAATG AAATGAGAAC 13200

AGGACAAATC GATTTCTAAC AATGTTTTAG AAGTAGAGGT GTACTATTCT AGTTTCAATA 13260

TACTGTAATA TGTGATGAAA ATGCCAGTAA TGATACCGAG AAAAAAGCTG AGAAACTTTT 13320

CCCAGCTTTA TTTGTTATAG TCAAAGAGAA TGACTTGTTC CTGTGCATCT ACATGAGCAT 13380

GGACCCCAAA GGGTACAATT GCTCTTGGAG TTGCGTGGCC GACATTCAGA TTATAGACAA 13440

TCGGGATATT GCTGTCAATG ATATCCAATA GTGCCTCTTT ATAGTCGTCA TGGAAAGTTT 13500

CATCCATAGG TTTTCCGACC AAGAGTCCAT TGATGACCGC GAATATGCCA GTGTCCTTTA 13560

AAGTTAGCAA CATCTTTTTG AAGTCTTCTG GCTTAGGCTT TTCTTCGCTT GTTTCGAGCA 13620

AGAGGATTTT CCCTTCCGAG TGTGACAAGT CAGGGAAAAG TTTGTATTTT TGGCAGAGTT 13680

CCGTGCTATC TGCGTATCGA GAGTTGTCAA AGATATCGTA GAGGGATTCG AGGCAACCAC 13740

CGAGGATTTT CCCCTCGAAC TGGGCACTTC CTTGCAACAA GTCAAAACCT GTATTTGTAT 13800

GACTGACACG AGGTGTTCCC AGGGCCGTGG GACTAAAATC AGTTCGTTCC TCATACCAAA 13860

CGTCACTAGG GCGGATTTCT GAAATTCTTC CCGTCTCAAT CAATTCTTTA AAGTAGTGAA 13920

GGCTATAGGC TAGCATTTCT TTGTCTAATT CACAAATGTC TGCTAAAAAG GATTGACCAT 13980

AAAAAGTCTT GATTCCTAAT TTATGCAACA TGAGGTGGTT CATGGTTGTA TCCGAGAAGC 14040

CAAGAAAAAT TTTTTGCTTG ATAACCTTTT GGAGTTGGTC ATTTTCAAAA AGATAAGGTA 14100

GCAAGCGATA GGTATCGTCT CCACCGATGG CACATAGGAT CATGTCGATG CTATCATCAG 14160

AAAAGGCATG AATCAAATCC TCTGCACGAG CTTCAGGATG GTCCTTGATA AAGTCTAATC 14220

CTTTTAACGA ATGGGGCAAA AAGATGGGAT TGGTCCCAGA TCCTTGAGAC GTT 14273

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9828 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GTGAAGTGCG GCAAAAGGTG CAAGTGATGA GCTCAGGTTC TTTAGCTCTT GACATTGCCC 60

TTGGCTCAGG TGGTTATCCT AAGGGACGTA TCATCGAAAT CTATGGCCCA GAGTCATCTG 120

GTAAGACAAC GGTTGCCCTT CATGCAGTTG CACAAGCGCA AAAAGAAGGT GGGATTGCTG 180

CCTTTATCGA TGCGGAACAT GCCCTTGATC CAGCTTATGC TGCGGCCCTT GGTGTCAATA 240

TTGACGAATT GCTCTTGTCT CAACCAGACT CAGGAGAGCA AGGTCTTGAG ATTGCGGGAA 300

AATTGATTGA CTCAGGTGCA GTTGATCTTG TCGTAGTCGA CTCAGTTGCT GCCCTTGTTC 360

CTCGTGCGGA AATTGATGGA GATATCGGAG ATAGCCATGT TGGTTTGCAG GCTCGTATGA 420

TGAGCCAGGC CATGCGTAAA CTTGGCGCCT CTATCAATAA AACCAAAACA ATTGCCATTT 480

TTATCAACCA ATTGCGTGAA AAAGTTGGAG TGATGTTTGG AAATCCAGAA ACAACACCGG 540

GCGGACGTGC TTTGAAATTC TATGCTTCAG TCCGCTTGGA TGTTCGTGGT AATACACAAA 600

TTAAGGGAAC TGGTGACCAA AAAGAAACCA ATGTCGGTAA AGAAACTAAG ATTAAGGTTG 660

TAAAAAATAA GGTAGCTCCA CCGTTTAAGG AAGCCGTAGT TGAAATTATG TACGGAGAAG 720

GAATTTCTAA GACTGGTGAG CTTTTGAAGA TTGCAAGCGA TTTGGATATT ATCAAAAAAG 780

CAGGGGCTTG GTATTCTTAC AAAGATGAAA AAATTGGGCA AGGTTCTGAG AATGCTAAGA 840

AATACTTGGC AGAGCACCCA GAAATCTTTG ATGAAATTGA TAAGCAAGTC CGTTCTAAAT 900

TTGGCTTGAT TGATGGAGAA GAAGTTTCAG AACAAGATAC TGAAAACAAA AAAGATGAGC 960

CAAAGAAAGA AGAAGCAGTG AATGAAGAAG TTCCGCTTGA CTTAGGCGAT GAACTTGAAA 1020

TCGAAATTGA AGAATAAGCT GTTAAAGCAG TGGAGAAATC CGCTACTTTT TCGATTTTTG 1080

ATTCAAGTTT TTAGATTATA TATAGTAGCT TGAAATAAGA TATGAACAAC TCTATTAGGA 1140

AAGTCAAATT AATTTCTAGA AATGTTTTAG CAGCTACAGC GTACTATTCC AAACTCAACC 1200

AACTATAATA GATCGAAACT AGAATAGTAC ATATCTACTT CTAAAACATT GTTAAAAATC 1260

GATTTGACTT TCCTTATTTC ATTCCGCTAT ATATAGTTTG CTGTTTCTTG TCGCTCCTCT 1320

GGAAAGCTGA TATAATAGCT TTATGAATAA AAAACGAACA GTGGACCTGA TACATGGTCC 1380

GATTCTTCCC TCGCTCTTAA GCTTCACCTT TCCAATTTTG CTATCAAATA TTTTTCAACA 1440

GCTCTATAAC ACTGCTGATG TCTTGATTGT TGGACGATTT CTTGGTCAAG AATCCTTGGC 1500

TGCAGTAGGA GCGACGACAG CGATTTTTGA CCTGATTGTA GGTTTTACAC TTGGTGTTGG 1560

CAATGGCATG GGGATTGTCA TTGCTCGTTA TTATGGGGCT CGGAATTTCA CTAAAATCAA 1620

GGAAGCAGTA GCAGCCACCT GGATTTTAGG TGCTCTTTTG AGCATTCTAG TTATGTTGCT 1680

GGGCTTTCTT GGCTTGTATC CTCTCTTGCA ATACTTAGAT ACTCCTGCAG AAATTCTTCC 1740

TCAATCTTAT CAATATATTT CTATGATTGT GACCTGTGTA GGTGTCAGCT TTGCTTATAA 1800

TCTTTTTGCA GGCTTGTTGC GGTCTATTGG TGACAGTCTA GCAGCCCTGG GATTTCTGAT 1860

TTTCTCTGCC TTGGTTAATG TGGTTCTGGA TCTCTATTTT ATTACGCAAT TGCATCTGGG 1920

AGTTCAATCC GCAGGACTTG CTACCATTAT TTCGCAAGGT TTATCAGCGG TTCTCTGCTT 1980

TTATTATATT CGTAAAAGTG TGCCAGAACT CTTGCCACAG TTTAAACATT TCAAATGGGA 2040

CAAAAGCTTG TACGCGGATC TCTTGGAGCA AGGTTTGGCT ATGGGCTTGA TGAGTTCAAT 2100

TGTATCTATC GGCAGTGTGA TTTTACAGTT TTCTGTTAAT ACATTTGGTG CAGTGATTAT 2160

TAGTGCCCAG ACGGCAGCTC GACGCATTAT GACCTTTGCC CTTCTTCCTA TGACCGCTAT 2220

TTCTGCATCA ATGACGACCT TTGCTTCTCA GAATCTAGGA GCTAAGCGAC CTGACCGTAT 2280

TGTTCAAGGT CTTCGAATCG GCAGTCGTTT AAGTATATCC TGGGCAGTTT TTGTTTGTAT 2340

TTTCCTCTTT TTTGCCAGTC CAGCTTTGGT TTCCTTCTTG GCTAGTTCGA CAGATGGTTA 2400

CTTGATAGAA AATGGAAGTC TCTATCTGCA AATCAGTTCA ACCTTTTATC CCATTTTGAG 2460

CCTCTTGTTG ATTTATCGCA ATTGCTTGCA GGGCTTGGGG CAAAAGATCC TTCCTCTAGT 2520

TTCTAGCTTT ATTGAACTAA TCGGAAAAAT CGTTTTTGTG GTTCTGATTA TTCGTTGGGC 2580

AGGATATAAG GGTGTTATCC TTTGTGAACC TCTTATCTGG GTTGCCATGA CAGTTCAACT 2640

GTACTTCTCA TTATTCCGTC ATCCCTTGAT AAAAGAAGGC AAGGCAATCT TGGCAACCAA 2700

AGTGCAATCC TAGTTGGATT TACTGAATAA AATCCATTTC CTCTAGTGAA AATCGAAAAA 2760

ACTTGTGTTC TCTTCTTTAG TTTGGTGTTG AAAATAGTTT AACAGACTTT TGACTTCTTT 2820

TATATGATAT AATAAAGTAT AGTATTTATG AAAAGGACAT ATAGAGACTG TAAAAATATA 2880

CTTTTGAAAA TCTTTTTAGT CTGGGGTGTT ATTGTAGATA GAATGCAGAC CTTGTCAGTC 2940

CTATTTACAG TGTCAAAATA GTGCGTTTTG AAGTTCTATC TACAAGCCTA ATCGTGACTA 3000

AGATTGTCTT CTTTGTAAGG TAGAAATAAA GGAGTTTCTG GTTCTGGATT GTAAAAAATG 3060

AGTTGTTTTA ATTGATAAGG AGTAGAATAT GGAAATTAAT GTGAGTAAAT TAAGAACAGA 3120

400

TTTGCCTCAA GTCGGCGTGC AACCATATAG GCAAGTACAC GCACACTCAA CTGGGAATCC 3180

GCATTCAACC GTACAGAATG AAGCGGATTA TCACTGGCGG AAAGACCCAG AATTAGGTTT 3240

TTTCTCGCAC ATTGTTGGGA ACGGTTGCAT CATGCAGGTA GGACCTGTTG ATAATGGTGC 3300

CTGGGACGTT GGGGGCGGTT GGAATGCTGA GACCTATGCA GCGGTTGAAC TGATTGAAAG 3360

CCATTCAACC AAAGAAGAGT TCATGACGGA CTACCGCCTT TATATCGAAC TCTTACGCAA 3420

TCTAGCAGAT GAAGCAGGTT TGCCGAAAAC GCTTGATACA GGGAGTTTAG CTGGAATTAA 3480

AACGCACGAG TATTGCACGA ATAACCAACC AAACAACCAC TCAGACCACG TTGACCCTTA 3540

TCCATATCTT GCTAAATGGG GCATTAGCCG TGAGCAGTTT AAGCATGATA TTGAGAACGG 3600

CTTGACGATT GAAACAGGCT GGCAGAAGAA TGACACTGGC TACTGGTACG TACATTCAGA 3660

CGGCTCTTAT CCAAAAGACA AGTTTGAGAA AATCAATGGC ACTTGGTACT ACTTTGACAG 3720

TTCAGGCTAT ATGCTTGCAG ACCGCTGGAG GAAGCACACA GACGGCAACT GGTACTGGTT 3780

CGACAACTCA GGCGAAATGG CTACAGGCTG GAAGAAAATC GCTGATAAGT GGTACTATTT 3840

CAACGAAGAA GGTGCCATGA AGACAGGCTG GGTCAAGTAC AAGGACACTT GGTACTACTT 3900

AGACGCTAAA GAAGGCGCCA TGGTATCAAA TGCCTTTATC CAGTCAGCGG ACGGAACAGG 3960

CTGGTACTAC CTCAAACCAG ACGGAACACT GGCAGACAAG CCAGAATTCA CAGTAGAGCC 4020

AGATGGCTTG ATTACAGTAA AATAATAATG GAATGTCTTT CAAATCAGAA CAGCGCATAT 4080

TATTAGGTCT TGAAAAAGCT TAATAGTATG CGTTTTCTTG TGGAGATATT TCCTTCAATT 4140

TTGCTACTAT ATTAAACAAA AATCAAAAAG CAAACTAGAA AGTTATGCTC AAATAAAATC 4200

TAAATTTGAC AATGTAAACC GAGTCGGATA GCTTTAAGTA CTGTTTTGAG GTTGAAGATA 4260

CGATTTTTGA TAGGAACTCA TCAATTTTAG ATTTTTAAGC AGCATCAATA AATTGCTTCC 4320

TTGTTTTGTC ATAATTTTTT TATTTAAAAA ATTATGACma GAGTGTGCTA TTCTTTTTAT 4380

GAGAGGTGTA TGAATATGAT AAATGTATGT GATAAATGTA TGTGATGTTG GAAAAAGAAT 4440

AAAAGAACTT AGAATATCTT CAAATCTTAC TCAAGATAAG ATTGCTGAGT ATTTGTCTTT 4500

GAATCAAAGC ATGATTGCCA AAATGGAAAA AGGTGAAAGG AATATCACGA ATGGATTTAA 4560

GTAATAAAGC TTCAAATCTT AGAAAAAAGT TGGGAGCTGA TGGTGAATCG CCGATAGATA 4620

TTTTTAAATT GGTACAAAAG ATAGAAAATT TGACGCTGGT ATTTTATGGA CTCGGAAAGA 4680

ATATTAGCGG AGTCTGTTAT AAAGGAACTC AGTTCAGTCT CATTGCAGTC AATTCAGACA 4740

TGCCATTAGG AAGGTAAAGA TTTTCTTTAG CACATGGACT GTATCATCTT TATTATGATG 4800

AGGTGAAGAA GAGTTCAGTC AGTCTTATCT TGATTGGTGA AGGAGATGAA ACTGAAAGAA 4860

AAGCGGATCA GTTTGCTTCT TATTTTTTAA TTTTCCCATC TTCACTGTAT AGGATGGTTG 4920

AGGAAATCAG AGAAAATGCC AATAGAACTC ATCTTGAAGT AGAAGATATT ATAAAATTGG 4980

GTCAGTTTTA TGGTATCAGT CATAAAGCTA TGTTATATAG ATTGAGGAAT GATGGATACC 5040

TTGATGCAGA AGAAATTAAA AATATGGATA TTAGTGTTAT AGAGACAGCT TCAAGATTAG 5100

GCTATGATAC AAGTTTATAT CGTCCTTTGT CAGAAAGTAA AAAAGAAATG GCATTAGGAT 5160

AATATATTAA TTCAACTGAA CAACTTTTAG AAAATAACAG AATTTCGCAA GGGAAGTATG 5220

AGGAACTGTT ACTAGATGCT TTCAGATATG ATATTGTATA TGGGCTAGAT GAAGAGGGGG 5280

GAGTTGTCGT TTGACTAGTC GTGTATTTAT TGATGCAGAT TGTATTTCAG TATTTTTATG 5340

GGTTGGCACT GAACATCTTT TAGAAAAGCT CTATTTGGGT AAAATTGTTA TTCCACAAGA 5400

GGTGTATGAT GAAATCAATA TACCTACAAT TCCCCATTTA AAATCTAGGA TAGATCAGTT 5460

GGTAGCTAAG GGTTCAGCTG AGATTGTGAG CATAGACATT GGAACTGAAG AATACGCATT 5520

ATATAGAGAT TTAACAAGAA ATCATGATAG TAACAAGATT ATTGGTAAGG GAGAAGGGGC 5580

ATCTATTTCC TTAGCGAAAA AGCATAATGG GATATTAGGA AGTAATAACC TAAGAGATGT 5640

TAAATCATAT GTAGAAGAAT TTTCTTTAGA ATATATGACA ACAGGAGATA TACTGATTGA 5700

AGCGTTTAAA GCGTAATTTA TTACTGAATA AGAGGGCAAT CATATCTGGA ATAATATGCT 5760

TAAAAAGAGA AGGAAAATTG GTGCAAATTC ATTTTCAGAC TATCTTCGTG GAAGTATTCA 5820

TCAAAATAGA CAAAAATAAA TTTGGATAAA TCGAACTCAC TATTCAGGAG GCATATGAGC 5880

AATTCGAAAA AGAAAAGTGT CAAATTGAGC CTATAGGAGT AGAAGTGAAA TAGTAAGTCC 5940

TGCATAGTGG ATGAGAGAAA AGTTCTCCTT GAAGTTTTCC TGAACTATCA GTCGCATGTC 6000

AAACGATATG TAGGGTAATG TGAGAGGGGA TAGCGAGTAG TTTTTGGTTA TTTTATCAAA 6060

AAACTTATAT TTTATTATAC CGAATGATAA AATATAATAA AAATGATAGA ATAAGGAAAA 6120

AACATGAATG TCAAAAAGAT AATGTCAATT TTTCAATCCT TTTATGTTGA TGTCAGTATT 6180

GAGGAACTGA CTTTGACTTT ACCAATCAGT TTTGTAAAAA GGTTTGAGTA TACTCAAATG 6240

ACTTTTCATA AGGAATCATT TTTATTGATT AAAGAAAAGA GAAGGGGGAG TTTGAGTTCA 6300

TTTGTTACTC AGGCTCGCAC TATGGGTGAA AAAGCCAATA TGGATGTTGT TTTGGTGTTT 6360

TCGAAGTTAT CAGACAGTGA AAAAAAGCAA TTACTTCAAG CTAGAGTTCC GTTTGTAGAC 6420

TTTAAGGGAA ACCTCTTCTT CCCTCCATTG GGACTAGTAC TCAATGCGAA TGATACTGAA 6480

GTCCCTAAGG AATTAACACC TAGCGAACAA TTAACGTGGA TTGCCTTTTT ATTGACAAAA 6540

GGTCAAAAAG TAGTAGATGT TGATTTGCTT TCACAAGTCA CTGGACTTCC AAACTCAACA 6600

ATTTATAGGT GTTTGAGGAC TTTTAAAGCT TTATATTGGT TAAACAAGCA AAATAAGCTT 6660

TACACATATA CGGTGTCAAA GAAAGAATTA TTCTTAAAAT CCGTGTCATG TTTATTTAAT 6720

CCCATCAAAA AACGGATTTT ATTGCCAGAT GGCGATATAA AGCAGATAAA ATCTGTTTCT 6780

AACCTTCTAT ATGGTGGTGC TTATGCTTTG TCGCATTCAA CTTTTTTAGC TGAAACGGAT 6840

GAAAATATTA GCTATGTCAT ATGGCAGAGA AAATTCAATC AGTTATCCTT GCCACTTTCT 6900

CAGCATGTTT TAAAATGAAA GATGCTAGAG ATATGGAAAT ATCGTCCTTT TGTATCTGAG 6960

TTTTGGAATG ATTTTAAAAA TAATCATGAT AAACAATTTG TAGATCCGAT TTCTCTTTAT 7020

TTGACCTTAA AAGATGATGA TGACCCACGT ATAGAGGAAG AGAGTGAAGC ACTAGAAAAT 7080

ATGATATTAC AGTATCTGGG AGAAGATGAT GCCAGCTAAT ACGAAAGTTA TTTTTCAAGA 7140

AATGTTTGCG GATTTTCAGA ACTATTATGT TCTGATTGGG GGAACTGCTA CCTCTATCGT 7200

ATTGGATTCG CAAGGATTTA AAAGTCGCAC AACAAAAGAT TATGATATGG TCATCATTGA 7260

TGAAGTAAAA AATAAGGAAT TTTATACTAC CTTGAATCAT TTTTTAGAAT TGGGAGAGTA 7320

TCAAGGAAGT CAGAAAGATG AGAAAGCGCA GCTTTTTCGA TTTACAACAA CTAATCCTGA 7380

GTTTCCTTCT ATGATTGAAC TATTTAGTAT CTTACCAGAA TATCCATTAA AGAAGGACGG 7440

TCGAGAAATT CCCTTACATT TTGACCAAGA TGCTAGTTTA TCAGCCTTAT TATTGGATGA 7500

AGATTATTAT AATATATTGG TGCATGAAAA AGAAACCATT CAGGGGTATT CGGTATTGAG 7560

TAATTGTGGT TTATACTCTT CGAAAATCTC TTCAAACCAC GTCAGCTTCC ATCTACAACC 7620

TCAAAACAGT GTTTTGAGCA GCCTGCAGCT AGCTTCCTAG TTTGCTCTTT GATTTTCATT 7680

GAGTATTAAT TATTTTTAAG GCTAAAGCTT GGCTGGATAT GAGGGAGCGC TCTGCCACAG 7740

GTGCTCAAGG TTTAAGTAAG TCCATTAAAA AGCATTTGAA TGACCTTACC CGTTTGACAG 7800

CTTCCTTGCT AGGAGATGAA AAGTTATCGG CTATAACATC AAGTAGTGCG GTAAAAGCAG 7860

ACATGCACCG CTTTGTGATA GAATTAGAGC CTGTGAAGTC AACTATTCTT CAAAATAATG 7920

ACATTTCATT GGATCAAAAT GAAATTTTTG AAATTCTGAA AAATTTTCTC GATGGTTAAA 7980

ATAATTGTAG CGAGATGGCT ATATTGAATT CGTCTATATC TGGAAACTAG AAAAAACTTC 8040

AATTTCAGGA GAAAATGAAG TCAATCTTCC CACAATCAAA CGTATAGTAT CAAGGTTTTT 8100

CAAGACCTGA TATTATGCGT TTTTTGCTTT TCAAAACTTT TTGCCCAGTC TTCGTTTTTA 8160

TCCTCTAGTC ACTTGATTTG TTTCAGGTGG TTTTTTAGTA TAGTAGAATG AAACGAGAAC 8220

AGGACAAATT GATCAGGACA GTCAAATCGA TTTCTAACAA TGTTTTAGAA GCAGAAGTGT 8280

ACTATTCTAG TTTCAATCTA CTATAGTTAA ATCTGCGGTC AAGTCTACTG GTGAATCTAT 8340

GATTGTAATA CTCTTCCAAA ATCTCATCAA CCACGTCAGT CTTGCCTTGC AGTCTGTATC 8400

TTACTGACCA AGCTAGTGAT GGATTTAGAA TAGGTGATTT GGAGCGTCCT ATTAGCTAGG 8460

AAATGCTGCT CATAGTCCTT TGCTGAGGCT AGGGTGTTTC AACATTCAAC ACTCAACTGG 8520

TTGATCTAGT TGATAGGAAG GGAGTTACTA TAAAATACTC AGGCTTCCAT CATATTTTTT 8580

GAAACGATTG TGTAATCAAA ATGTACCAAT ATTGTAGTAT TGGTACAGAA GATGTTGTGA 8640

ATGGATAAAT ATATCATAAC TGCTATCTCA AAAAGATTTC ATATGTCTGT GCATATATAA 8700

TAGACTTCCT GCAAAACTAG AATCCTAGTT CATGATTGAT AATACCAGCA ATCAAATTCA 8760

TTCGTAATCC AAAGCGTTTA CGATGATTTC GATAGGTTGT TGAAAACATT TTAAACGTTT 8820

CTACTTTGGC AAAGATGTTC TCAACCTTGC TTCTCTCCTT AGATAGCGCA TGGTTATAGG 8880

CTTTATCTTC AGCTGTTAGC GGCTTGAGTT TGCTGGATTT ACGTGGAGTT TGTGCTTGAG 8940

GACATATCTT CATGAGCCCT TGATAACCAC TGTCAGCCAA GATTTTACCA GCTTGTCCGA 9000

TATTTCTGCA ACTCATTTTG AACAACTTCA TATCATGACA ATAGTTCACA GTGATATCCA 9060

AAGAAACAAT TCTCCCTTGA CTTGTGACAA TCGCTTGAGC CTTCATAGCG TGAAATTTCT 9120

TTTTACCAGA ATCATTCGCT AATTCTTTTT TTAGGGCGAT TGATTTTTAC TTCCGTCGCA 9180

TCAATCATTA CCGTGTCCTC AGAACTAAGA GGAGTTCTTG AAATCGTAAC ACCACTTTGA 9240

ACAAGAGTTA CTTCAACCCA TTGGCTCCGA CGGATTAAGT TGCTTTCGTG AATACCAAAA 9300

TCAGCCGCAA TTTCTTCATA AGTGCGGTAT TCTAGGCTTA ATTTAGGTTT TCGTCCACCT 9360

TTTGCGTGTT TAAGTTGATA AGCTGTTTTT AATACAGCTA ACATCTCTTT AAAAGTCGTG 9420

CGCTGAACAC CAACAAGACG CTTAAATCGT GTATCAGTTA ATTGTTTACT TGCTTCATAA 9480

TTTCGCAGGG AGTCTATTGA CTCTTTGGTA GGTGTCAATG TTTTTTTCAT CTATCCCGAG 9540

AATTATTTTC CCGCCATTTG TATTTGCAAA TGCTGAGTAG GTTTCCCAGA AAGACTCTGG 9600

AAGATTGTTT TTAGCTTTTT TGTATTCTAA ATCAACCCCT TCAAATTTTA AGTCCATATT 9660

TTTCCTTTAC ATCTGTTTTT TGTGGTTCTG GTATTTGTTC AAGTTGAGTG ATAATATAGC 9720

GAATTGAATT TCGAGAGTTT TTACTCAGTT AATTTCTTTT TTAACCCACT TTAATTGCTT 9780

TTTTAACACG GGTTAAAAAA GAAATTAAAG TGGGTTAATT TTTCTTGA 9828

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3369 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

CCGCGAAAGA TATTTTTGAA CAAGAGTTTG GACGTGAGGT CGGTGCTTAT AATAAAGTAG 60

AAGTTGACGA GTTTTTAGAC GATGTCATCA AGGACTATGA AACCTATGCT GCCTTGGTCA 120

AGTCACTTCG TCAGGAAATT GCGGATTTGA AGGAAGAATT AACTCGTAAA CCGAAACCTT 180

CACCAGTTCA AGCAGAACCC CTTGAAGCGG CAATTACAAG TTCTATGACG AATTTTGATA 240

TTTTGAAACG CCTGAATAGA TTGGAAAAAG AAGTTTTTGG TAAACAAATT TTAGATAACT 300

CAGATTTTTA AGTAGTTATT TGAGATGTGC AATTTTTGGA TAATCGCGTG AGGAGAATTG 360

TTTCTCATGA GGAAAGTCCA TGCTAGCACA GGCTGTGATG CCTGTAGTGT TTGTGCTAGG 420

CGAAACCATA AGCCTAGGGA CGAGAAATCG TTACGGCAGT TGAAATGGCT AAGTCCTTGG 480

ATAGGCCAGA GTAGGCTTGA AAGTGCCACA GTGACGGAGT CTTTCTGGAA ACAGAGAGAG 540

TGGAACGCGG TAAACCCCTC AAGCTAGCAA CCCAAATTTT GGTCGGGGCA TGGAGTACGC 600

GGAAACGAAC GTAGTATTCT GACTGCTATC AGCTAGAGCT GTTAGTGGTA GACAGATGAT 660

TATCGAAGGA AGTGGTCCTA GTCACTTCTG GAACAAAACA TGGCTTATAG AAAATTGCAT 720

ATAGGTTGGG GCTGAGAAAT TTTCTCAACC TCATTTTTTA AAGTGGACAT ATAGAAAGGT 780

CTTGCAAGAC TGTAACATGA AAAAAGAATT TAATTTAATT GCAACTGTGG CAGCAGGGCT 840

TGAGGCTGTC GTTGGTCGTG AAGTGCGAGA GTTGGGCTAC GATTGTCAGG TTGAAAATGG 900

ACGTGTTCGT TTTCAAGGAG ACGTGAGAGC TATTATCGAA ACCAACCTTT GGCTTCGGGC 960

AGCAGATCGT ATCAAAATTA TCGTAGGAAC GTTCCCAGCT AAGACTTTTG AAGAGCTATT 1020

TCAGGGAGTT TTCGCTTTGG ATTGGGAAAA TTATTTACCA CTTGGAGCTC GGTTCCCGAT 1080

TTCAAAAGCT AAATGTGTTA AGTCCAAACT TCACAATGAG CCCAGTGTTC AGGCTATTTC 1140

TAAGAAAGCT GTTGTCAAGA AATTGCAGAA ACACTATGCT CGCCCAGAAG GGGTTCCTCT 1200

GATGGAGAAT GGCCCAGAGT TTAAGATTGA GGTCTCTATT CTCAAAGATG TGGCAACTGT 1260

CATGATTGAT ACGACCGGGT CTAGCCTCTT TAAACGTGGT TATCGTACCG AAAAAGGTGG 1320

CGCTCCTATC AAGGAAAATA TGGCAGCAGC CATTTTACAA CTTTCTAACT GGTATCCACA 1380

CAAGCCTTTG ATTGATCCGA CCTGTGGTTC GGGGACTTTC TGTATTGAGG CAGTTATGAT 1440

TGCTAGAAAG ATGGCGCCAG GTCTTCGTCG CTCTTTTGCA TTTGAGGAAT GGAACTGGAT 1500

CAGCGATCGC TTGATTCAAG AAGTGCGCAC AGAAGCGGCT AAAAAAGTAG ACCGTGAGCT 1560

TGAGCTGGAT ATCATGGGCT GTGATATTGA TGCTCGCATG GTGGAAATTG CTAAGGCCAA 1620

TGCTCAGGTA GCTGGTGTTG CAGGAGACAT TACTTTTAAG CAGATGCGCG TGCAGGATTT 1680

ACGTTCCGAT AAAATCAATG GAGTAATCAT TTCCAATCCG CCTTATGGTG AACGTTTGTC 1740

AGATGATGCA GGGGTGACCA AGCTCTATGC TGAGATGGGG CAAGTATTTG CACCGCTGAA 1800

AACTTGGAGC AAATTTATCC TGACTAGTGA TGAAGCTTTT GAAAGCAAGT ATGGTAGCCA 1860

AGCAGATAAG AAGCGTAAGT TATACAACGG AACCTTGAAA GTGGATCTAT ATCAATATTT 1920

TGGTCAGCGT GTCAAACGGC AAGAGGTAAA ATAGAAAGGG ATACTCATGA GTAAAAAAAG 1980

ACGAAATCGT CATAAAAAAG AAGGTCAAGA ACCGCAATTT GATTTTGATG AAGCAAAAGA 2040

GCTAACAGTT GGTCAAGCTA TTCGTAAAAA TGAAGAAGTG GAATCAGGAG TCTTGCCTGA 2100

GGATTCCATT TTGGACAAGT ATGTTAAGCA ACACAGAGAT GAAATTGAGG CGGATAAGTT 2160

TGCGACTCGT CAATACAAAA AAGAGGAGTT CGTTGAAACT CAGAGTCTGG ATGATTTAAT 2220

TCAAGAGATG CGTGAGGCTG TAGAGAAGTC AGAAGCTTCT TCGGAGGAAG TTCCATCTTC 2280

TGAAGACATC TTACTACCCT TGCCTCTGGA CGATGAGGAG CAAGGCTTGG ATCCTCTATT 2340

GCTAGATGAT GAAAATCCAA CAGAAATGAC TGAAGAAGTG GAAGAGGAGC AAAACCTTTC 2400

TCGTCTGGAT CAAGAGGACT CAGAAAAGAA AAGTAAAAAA GGCTTTATTT TGACCGTTTT 2460

GGCGCTTGTA TCAGTAATTA TTTGTGTCAG TGCTTATTAT GTCTACCGTC AAGTGGCTCG 2520

TTCGACTAAG GAAATTGAAA CTTCTCAATC AACTACAGCC AATCAATCGG ATGTGGATGA 2580

TTTTAATACA CTTTATGACG CCTTTTACAC AGATAGCAAT AAAACGGCTT TGAAAAATAG 2640

CCAGTTTGAT AAACTGAGTC AACTCAAGAC TTTACTTGAT AAGCTGGAAG GTAGTCGTGA 2700

ACATACGCTT GCCAAATCTA AATATGATAG TCTAGCAACG CAAATCAAGG CTATTCAAGA 2760

TGTCAATGCT CAATTTGAGA AACCAGCTAT TGTGGATGGT GTGTTGGATA CCAATGCCAA 2820

AGCCAAATCG GATGCTAAAT TTACGGATAT TAAAACTGGA AATACGGAGC TTGATAAAGT 2880

GCTAGATAAG GCTATCAGTC TTGGTAAGAG CCAGCAAACA AGTACTTCTA GCTCAAGTTC 2940

AAGTCAAACT AGCAGCTCAA GTTCAAGTCA AGCAAGTTCA AATACGACTA GTGAGCCAAA 3000

ACCAAGTAGT TCAAATGAGA CTAGAAGTAG TCGCAGTGAA GTCAATATGG GTCTCTCGAG 3060

TGCAGGGGTT GCTGTTCAAA GAAGTGCCAG TCGTGTTGCC TATAATCAGT CTGCTATTGA 3120

TGATAGTAAT AACTCTGCCT GGGATTTTGC GGATGGTGTC TTGGAACAAA TTCTAGCGAC 3180

TTCACGTTCA CGTGGCTATA TCACTGGAGA CCAATATATC CTTGAACGTG TCAATATCGT 3240

TAACGGCAAT GGTTATTACA ACCTCTACAA GCCAGATGGA ACCTATCTCT TTACCCTTAA 3300

CTGTAAGACA GGCTACTTTG TCGGAAATGG CGCTGGTCAT GCGGATGACT TAGATTACTA 3360

AGCAGTCGG 3369

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9713 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

AAGTTTACAA TTTAAATGAA TTAACAATTT TCCCAACTAA AAGCACTCCA GTTACCGCAA 60

CGTTTGTACT GAATGTACTA AATCGCATTC CATCAACTTC ATCTGTTTCG TCAACTTGAA 120

CAGATACTAA TTGAAGATTT AATACTTCTG CTGCCATAGC TAGCTCCTCC TATTTAAATT 180

TTTGGGATTA AGTACTTTAT CCACCCTCAT ATACTCTCTC CACCAGTAAA ATGCAAGCAA 240

TGATACAAAA TAGATTTAAC TATTTTATAT AGCGAAAACT TACAAATTTT TAAGAAATAA 300

TTTTTGCATT CTTAAAGATA AAATAGGAAC TTTTAGTAAT AAATATTAAA ATAAATAAAA 360

TAATAGATAC TATAAAATTT GGAAGTATTA ACCCCAAAAG ATTCATATCA TCTATTAAAA 420

TATCCTCTAA AGAGTAGTAT ATTAAAGCCA TAATTTTAAT GTTAAGTAAA AATGCAATTA 480

ATGAAGTAAC AAATGTCAAA AATATAGCCT CACCAACTTT AATCTTAACC ATCTGGTAAT 540

TAGAAGTTCC TAAAATTTCA AATTGCTGAA TCTCAATCCT TTCTTGATGC GATGACAAAA 600

ATGCAATTGA AATAATATTT GCAAGTACTA TCAAAATTGG TGCTCCTACA TAGACAATAA 660

ATGCTACTTT TAGCTCTAAA TCACTGTCAT CTTGAAATTG AGATAGTATA TTCTGAGAAA 720

TCATTTGAAA ACTAGAAATT AGTAATATAG CTCCTGTAAT TGCAGCACTG ATAGATTTTA 780

TATAAGACTT ACAATATAGT AAATTCCACT TCGAAACAAT GAACATAAAA TTATTTCTAA 840

ATATAATTAT AGAAAGTAGT TTGATAAAAC ATGACTGTAT AAAAGGAGAT AATTGATAAA 900

TAATCACAAT ATCTAAGATT ACAATATTGA ATATTATCTG GGCCTTCGCT AAAATTGTGC 960

TATCTTGGAA AATTTGTTGC AAAGAAAGCA ACCAGATAAC ACTAAAACCA GCCAATAGCA 1020

GTATTCTTTT TACTATTGAA AGAACATGCC TTATTTTAGA ACTCTTCCTA TTTCTAATCT 1080

TCTTGAACGT ATAAAAGCAA CCACTTAGAA AGGCTAAAAA TGAAATCAAC ACTACTGTAA 1140

TGATACATCC AACAGCACTC GTTTGAAATT GGATATCAGG TAATATATTT TCCCCGAAAA 1200

AGTATTGTAA AAAATAATAA TAATTTGACG TAACAAATAT AGAGCATAGA TATGCAATAA 1260

AACTAATAAT CGAGGAAATG ATAAAAATCT GTCCCCCCAC AAGAAATGAT AGTTGAAGGC 1320

GACTTGCTCC CAACACCTCC AGAAGTTCGT AATCATCTCT AAAAATTTCA ACCAACATAT 1380

TTATTATGTT AGAGAGCACA AAGAATAATG TTACTCCTCC GAATACTATC GGAAACATAA 1440

AAATTGGTTT AGGATCTGGA AGTCCGACAA ATACTTGCGA ATTATTCTCA ACATTAATTA 1500

CCCCATTAAC AGCCAATCCC ATAACTAAAC TCGAAACAAA AATTACTGGT GAAACGCCTA 1560

ACCATTGTTT CTTATTATGT AAAAATTGAT AGTAAACTAA TCTGAGCATC TCTATTCCTC 1620

CGTAGTTGAT TGTACCTCTA AGATTTTATA CAACTCTTCC CCGCTAGGTC TATGAAGTTC 1680

TTTGAAAATT TTTCCATCTT TCAATATTAA TGCACGATCA GTTTTCGAGG CCAATTCTAT 1740

ATCGTGCGTT ACCATAATTA CACACTTACC CGCCCCTACT AACTCTCTCA ATAATTCAAA 1800

AATTACTTCA CGAGAAACGC TGTCTAAAGC CCCAGTTGGC TCATCAGCAA ATATTATATC 1860

ACTATCAGCA ATAACCGCTC TAGCTATAGC AACCTTCTGT TGTTCTCCAC CAGACAGAGT 1920

TCCAACAAAA TCGTTTAAGC CAGCATTAAA CTTCATTCTT TTGAGTAAGT TTTCTACATT 1980

TTTAATAGTT AATTTTTTTT GTGATAATCG CAAAGGAAGT GCTATATTTT CTATTACCGG 2040

CAGGGAAGGT ATTAAATTGT ATGCTTGAAA TATAAAAGAT ACTTCGTTAC GTCTTATACT 2100

TGACAATTTT GCATTTCTGA TTTTATAGGG GTTGATTCCA TTTAAAATTA CTTCCCCACT 2160

TGTTGGTTCA AGCAAACTAG AAATACATTT TAATAAAGTT GACTTTCCAG AACCACTAAT 2220

TCCTAGAATA CTTATAAATT CTCCTCTCGA AGCAGAAAGA GAAACATTTT TCAGCACTTG 2280

CAACGTTTTA TTATTTCCTA GTAAAAATTG ATGATACAGC CCTTTCACTT TTAATATATA 2340

ATCTTTATCC ATATTCTTGC CTCCAATCAC TTAATTTTGA AAAGTGTTCC ATTTTCCAAT 2400

TTATATATAT CAGTGTATCT CTTGTCATTT AAGTCATAAT GATGTGAAAC TTCAATAAAT 2460

GAAATACCTA AATTGAACAG AATATCATGT ATGGAATTTG AATTATCATT ATCTAAATTA 2520

GCTGATATTT CGTCAAATAA GTACACTTTA TTATTTCTAA TCAGAGCTCT AGCTAAAGCT 2580

ATTTTTTGTT TTTGACCTCC AGACAAATTA CTACCATTTT CACCACATTG ATAATTTAGT 2640

ATATCTATCT TTTCTAATTC TTCATATAGA TTTACCTTTT TTAACACCTC AATTATCTGA 2700

TCATCTGAAA AATATTCATT TTGAAATAAA GTTACGTTCT CACGAATAGT AGTGTCAAAA 2760

ATATATGGTG TCTGATCAAC TGTTGGTATT GAATCTGAAC TCTTTTTCCC ATGTGATAAC 2820

AAATTTACAT AACCTTTTTG TGGCTTTAAA GAACCATTAA TTAAATTTAA AATCGTTGTT 2880

TTCCCACTAC CAGAAGTTCC TGTTAATAAT ACCCTAAATG GTGACTTAAA TGAGAAGTCA 2940

ATACTTAATT TATTTTCTGG TGTAATAGAA TATACAACAT CTTTCATGTG TATCTCATCT 3000

ATTGATGAAG TATACAGTCC GTTATTATCA TGTTCAGCGT CTATAAAATT CTTCTCTCCA 3060

CTTAAGTATT TTAAAAACGG TTTCCTTAAA TCTTTGGTTG TATTTATCTT ATTTAATGAA 3120

TAGGCAATTG ATTGTATCGG CCCTAAAACT TTATCGTTTG CTAAGAAAAT ACCTATCAGT 3180

TCACTAAAAG AAAGGCTTTT ATGATAAATT ACAAAATAAC ATCCTACAAC CAAGGGAACT 3240

AGAAAGCAAA AACCTGAAAT TAGTACTGCA ACCAATTTTG AAAGAACCTC TGATCGTTTC 3300

AAATTAAAAG TAGAATCTTC TAGTTTATCC AACTTTTTAT CCGACAAACT AATTATTTCT 3360

TTAGTAACAG AATAAGATTT TAATGTCTTA AAACCATTAA AAATTTCTTT TATTATGTGA 3420

GTATACTCTG CATTGCTGTT AGAGTACTCA TTAGCTGAAT TAGACAACAT CTTCTTCATA 3480

AAGACAGGTA CTATAATCGG CAATGCTGAT AATACAATAA ATATTATTGA nACTAGGAAG 3540

TTTAAATAAA GCATAAAACT TAGAGAGACG ATGAACAACA ATATTGAAGA AATTATTTCA 3600

AAAATTTGTC TAAAATAGTT TTCTTCGATT AATCTCAAAT CATTTGACAA AACTGAAATA 3660

ATAGATGAGT AATCTTTAAC CATTTCAGAA GAAAGATACT GTTCTCTAAA ATATCCTTGT 3720

TTAATTTTTA CATTTATATC TTTAGTTATT GATGCTTCCG TTACTTCTAA ATAGTAATTT 3780

GATATATAGA TTGCTGACCA ACCCAGAATA CTTATAGCAC CAAATCTTAG AACGTCAGAA 3840

AATGAGGAAG TCTGATTTAA ACTACCTGCA TATACAATAA TTCCTGAGAG CAAGACACCA 3900

TTAAACGAAG ATAGAAATAT TAAAATCCCC ATTAATATAA GTTTAGTCTT TTTTATAAAT 3960

TTTAAATAAT TCATAAGTTA TTCCTTCCCA CTTCTTCAAA GAAATAATTT AAAGTATCAA 4020

TCATTAAGAG AACATCTGAT GGAGTAAAAC CTCCATGACC AGCTGCTTTG TTTAAATACA 4080

ACAAACTTTT AACTCCAATA GAATTTAATT TCTTTGACCA CTCTATCACT TCGTTATTAT 4140

TAATATATGG GTCTTTCTCA CCCAAAATAT TAACTATAAC AGTATTTGAG TCTCGTGCCT 4200

TTTCAATATT TTGCATAGGC GAATATGACT TTATATAAGC CTTTACTTCA GGGTCTCTAA 4260

TATCTCCCCA CTCTGCTATT TCGGTCTTAG AAAGAGGATC ATTTGGATTC TGAAGTGTAT 4320

CATAAGGATT TATAAATGGC GAAAATAAGA GAATGCTTTG CAATAAATTT TTTTCCTCGT 4380

TCAACACCGC ACCAGCAATT ATTCCACCTG CACTAGAAGT TATTAAACCT AATCGCTTAC 4440

TGTCAATTAC ATCATTTTCC CTTAAATAAT TTACTCCCTC AATAAAATCT CTGATAGAAT 4500

TCCATTTGTT TAACGCCTTT CCTGAGCGAT ACCATTCACC ACCCAAATAG CCTCCACCTC 4560

TTACATGAAC TATAGCATAA ATAAAACCTG CATCTATTAT AGATAACATA ATTTCATCTA 4620

AATCAGAATT ATCATTCTTA CCATAAGCCC CATAGACACT TAGAATACAT TTTTTTCTTC 4680

TTGGGAGCTC ATCCGTATCT TCACTTTTCC AAAATAAAGA AATCGGTATG CTTACATCAT 4740

AACTGTCTTT TTTAGTCCAA ATCACCTTAG AAAAATATTT AGTATTATTC GATTTTATGA 4800

TGGGTCTTTC AAATTCAGTT TTTAATGTAT TTTCTATTAA ATCAAAACTA AGTATTTTTT 4860

CGTAAAAAGT TCTCCTCTCT AAAAACAGAA GAACACGATC AGAAAATGAA TTTTCATAAA 4920

GTGTTGTCTT TTCATCAAAT GTTATCTTAT TAACACTCAA CTCCCTCAAA CTATTATTTT 4980

TAAATGTAGC AAGATAAAAG ACGGAATTCG CTGCGTTTGA ACAGTCTAAA AGGATATAAC 5040

GTCCTATACA GTGAACTCTT CTAGCCCTAT CTTGATATGG TATAGTAATA GAAACTCTGT 5100

CTCCCGAAGA AGTTTCCCTT AGAATTAGTT GATCTTTCTT TTCTTCAGTT GAAGAGAGCC 5160

CAAGAAAGTA CTGTGCTTTT TCTGTACTAA ATAGAGCGAT ATCTCTAGGT GTTGGGGCTA 5220

CCGTTTCTGT GTAAGAGTGT CTAACAAAAC CCGTCCGGTC GAAACTGTAT AGAAAAATCC 5280

TGCCTTTCTG AAAGTCTACT GACTTTACAA AACAATTATT GCTATCAATG TGGACTATTT 5340

TTAATCGAAA AGAGCATTCG TTTTCTTCAA ACAGTTCCTC TTCTGTAAAG CTATCAAAAG 5400

ATTTATAGAA TAACTTACTT GGCCTCCCGT ACTCTTTGGA GCGAGTATAC ATAACACCGA 5460

ATTTACCCAA ATAGAACGAA CTTTCTACTG AAATATCTTC AATGATAAAT AACTCTTCCA 5520

TAGTATATTT TTTTATTCCA ATTAAATTAG TCGTACGCAG TGAGGATACA ACCAAAACTA 5580

TATAACTCTC ATCAGATGAA ATCCTAACAT CCTGTAAGAT ACTATCATCT GGCAAAGTAT 5640

ATTTTTCCAC ATCAAAGACA ATTTTAAGTG AATTTGAATT GTCTAAACTG GAAGAACTAA 5700

CCTTAGGAAT CCAGTCATTA TCTTCGACAT ACCATTCCTT TATTACACCA GTATTGGGTA 5760

TACTCCAATT ATCAAATTGG TACCAATATC GCCCTCTCCT AAATATCAAA GAATTCCATT 5820

TTTTTAATTC CTGAAATGAT GAAGAGATAG ACCTCTTATA GTGTGTTTTT TCCTGTATTG 5880

TATTTAAAAA TATTTCATTA CTCTGATTCA CAAGTATGAC CCCTTAATAA TGGTATCTAA 5940

ATATTATATT TGAGGAAGAA TCGTCAATTT ATTATCCATT ATTGATACCA ATCCAATTGC 6000

AACACCCGCA AATCCCGAAG CAATATCTGT TGTTATCTTT AAACCATTAT CTCCCGCAAT 6060

AACAAATCCT TCTTCAATTA CACACAAATA TCTATAAAGT TGTTCAATTA ATTTCTTTTG 6120

TCCTGAAAAG TTATCATCGA TATCACTATA TATATTATTA GCAACTTCAA GACCACAAAA 6180

TCCGTTAAAT AAACCTGGTA ATACACAAAA AACTACATCA GTTGCCCTCT CTAAAGAAGT 6240

TAAATATTTT AAGTATTTGC TTGACAAGAT TTCTTTATTT CTATTAATAA GTAAAAGCAG 6300

GCCAGCACTT CCAGTTGCTA GATATGGTAG TAATCTATGA CCTTGGCTGT ACTGCAATGA 6360

ATTATTACTA TCTACTTTAT AAGCAACTAA TTCTTTATCT ACAGCCAATT CTAGACCATT 6420

TTTATAGATA CTTTCACCAG TTAATTTATA AGCTTCACCG AAGAGCCAAG CTACCCCTGC 6480

GTGACCATAT AGTAATCCAC CAAAATTCTC ATAAGGATCG TTACTCTGAA CATCACTAGC 6540

GCCAACTTTA CAAAAAGTTT CTGGATTTTC TATATAATTT AAAGTATATT CTCTAAGCCT 6600

AATTAGTATT TCTTCTCCTA GTTTATTATC AATTCCCCCT TTACTAAGAA AATACAGTCC 6660

AACCAGTAAA ATTCCAGCCT GCCCACTATA TAAATTTTTA TTTTGTGAAT TCTCAAATAT 6720

CTCTATAAAA TGAGTTGTAA AAAGTTCAAC TGCCCGATCT ATCTCCCCAA ATTCATAAAT 6780

GAGCCAGATT GTACCAATTT TACCATCAAA AAGACCAGAA AGGGACGATT TCTTAAAATT 6840

ATTTACTGCC TCATTAATAA CCTGTGTTCG AATCTCATAA TAGTCATCAA ACTTGAAATT 6900

TTTTACTTTC TTAGCTAGTT GTTGATAACT CCAAAGGATA GCTAAATCTG AAAACGCAAT 6960

TCCTTGATTA AAATTCAGAC CATAATAATG AACTGGGAAG AATCTTGATT GAAATTCTTT 7020

ACGCCACTGT CCATAAGTTA GCGTAAACCC TCTCAATAAT TTTATAATAA AATCTTGTAT 7080

ATCTTGCTCA CTCTCGATAG TTCTAATCTC ATGCATGGGT TTTAAAACTT TTTTCCTGGA 7140

AATATTCTCA ATCTGTGGAC ATTTAGAATC TAGATATGAC AATAAACTTT CTACATAATC 7200

TATATGTTCT CTTGTATAAC CCAAAGACTC AAATAGTTTT TTTCCTTCTA TCCTGGTTTG 7260

ACTTACATAG TTGTATGTCA AATCCGATGT AGTTACTAGT GGCATGTATA AATAATGAGC 7320

TATTTGTCTA ATACCATACC AATCTATCTC ACTGGGAAGT GTTTCTCGCC ATGCTCTAAA 7380

ACCAGGGGCT GCAACTTTAT GTACAACTTT TTCATCATTT GAAAAGACAG CCTGTTCCCA 7440

GTCTATTATA CTAATCTCAT CTTCATCCTT AACCAAGATA TTTCCTAAAT GTAAATCTTG 7500

ATGATATACA TTTTCAGAAT GAAACTTATT CGTTAAATCG ATGAGTTTTT CTACTATCTT 7560

TGAAACTCTC AATAGATAAT CTTTGGTCTT ATCAACAACT TCATATAAAG GAAAATTATT 7620

GGTAACCCAT CTATTTAGTG GAACGCCCTT CATATGTTCA ATTCCTAAGA AGGTGTGCTC 7680

CCAGATCTTA CCGTGCCAGT ATATTTTAGG CGTCTCACTC CATTCATTTA GAATTTTTAG 7740

TGCTTTGCAC TCCGAAGCTA ATTTCTCTGA AGAATAAGTA CCATCAAATC CTAGACCTGT 7800

ATACGGTCTA GCCTCTTTTA AAATTATTTT TTTCCCATCT TCTTTTAGCC TAGCATTATA 7860

TATCCCACCA CTGTTTGAAA ATCTAATTGC ATTATCTATA ATAAAGGGAA AGTCTCCCTG 7920

TTTTTTATCT TTCTTGTCAA GCCATTTATT CAAAAAGTCA GGGGGCACTA TACCTTTTGG 7980

AATTTTAAAT ACTGGTAAAC GTTCATCTTT AACAACTTCA TCGCCAACAA TTAATTCATC 8040

AATAGCAACC TTCTTTTCAT CATCCCTTGA CGGCCTAAAC ACACCATACC TCAGATATAT 8100

TGGTGCTTCA TCCCAACGTT TATCGCTTAA AATATATGGC CCATTATATT GCTTTAAGGC 8160

ACTTTCTAAC CTTTGCAAAA CCGACTCTAA TTCATTTTGA TTTGGATAAC ATGTAATAAA 8220

TTTACCAGAA AATCCTCGAC TAACCAATTT CCCGTTTCGC ATGATAAATT TGTCTTCTGT 8280

ACTAAGATGT TTAAATGGAA TTCGCATTTC ATGGCAAATT TTTGCTACAT CTTGTAACAA 8340

TTCATGTGAA CTGTTATACT CTGAACTAAT GTGTATTTTC CACCCTTGTC TTTCAACAAA 8400

TTTTCCAATA GGGTATTGAT AAACCCACTC ATCATTATTC ATTACTTCGT GCCAATTAAA 8460

AGGCAGACTT ACTTGGTACT TTATGCTAGT ATCTGTACTA TAATCATTAT TAGTGAAAAA 8520

GAAAGGATGC TCCAAATTGA AATTATAATC CATAACAAAA TCTCCAAGAA ATTTTATCAA 8580

ACTTAATATA TCTATAGCTA GACAGACTTA TTTAAATAAA AAGGGAGAAT CCTTTGGATT 8640

CTCCCCATAT AAGCACTAAC ATTCCAACGT GCACATATTG GAACGACATC CATAACTCCA 8700

GAGAATCTCT AAAGTTTACA ATTTAAATGA ATTAACAATT TTCCCAACTA AAAGCACTCC 8760

AGTTACCGCA ACGATTTGTA CTGAATGTAC TAAATCGCAT TCCATCAACT TCATCTGTTT 8820

CGTCAACTTG AACAGATACT AATTGAAGAT TTAATACTTC TTCTGCCATA GCTAGCTCCT 8880

CCTATTTAAA TTTTTGGGAT TAAGTACTTT ATCCACCCTC ATTATACTCT CTCCACCAGT 8940

AAAATGCAAG CAATTATACA ATGTTGTCAC ATAGAAAATA ATGTTTCCGT AACTTTTCAA 9000

AGTAACTTCC ATCTCTCTCC CAAAACTGGA AGTTAGTTTT AGAAGTTACC TAAAAATCAG 9060

GTCACCTATT TTAAAAAAGC AGCAAACTAT AAACTAGTAG GTTCCACACC AAATGTAGTC 9120

CCATACTGCC CCATAAGTCA GATTTATAGC GCACCATACC TAAAAACATC CCAAGTGAAA 9180

CATACAAACA CCAAGCTAGA ATGGTTCCTG TATGATGTGC TAAGGCAAAT AAAACACTTG 9240

TCAAAGCAAC TCTGATATCT AATTTTCTGA CCAAATTCCA TAAAATTTCT CGATACAGAA 9300

ATTCTTCAAC CATACTCGCA TTGATTAAGA ACAATAAAAA TGAAAACCAA GGAATTTGAT 9360

GTTGAAGGCC AATTAAGTTT GCTTGATTCG TGCTTCCTTG AGCATGAATC AGACTAAAAC 9420

ATAGACTTAT AATCAGTAGG CTAACAAATT CAACACCAAG CCATTTCATC CTAGATTTCA 9480

TATTGACCTT ATGCGCTTGT TTGCGTTGGC CATACATCCA TAAAAAAGAA ATGAGTGACG 9540

AACCATAGAG AATCTGTAGT ATAGTTmACT CACCGATACA AAGAAATTTC AATAAGTATA 9600

GAGrTACCAA TAsGACATTT ACTTGTTGGA ATATATAAAC TGGAATTATT CTTTTCATAG 9660

TTACCTCCGA AATAAATCTT CATAATCTAA ATCTAATACC TGCACAATCC TTT 9713

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8657 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

AAAGAAATTG TCAGAGAGTG GCTAGATGAA GTAGCAGAGC GGGCTAAGGA CTATCCAGAG 60

TGGGTGGATG TTTTCGAGCG TTGCTACACC GATACCTTGG ACAATACGGT TGAAATCTTA 120

GAAGATGGTT CAACTTTTGT CTTGACTGGG GATATTCCTG CCATGTGGCT TCGAGATTCG 180

ACAGCCCAAC TCAGACCCTA CCTTCATGTA GCTAAAAGAG ATGCCCTCCT GCGTCAGACC 240

ATTGCAGGTT TGGTCAAACG TCAGATGACC TTGGTACTCA AGGATCCCTA TGCTAACTCC 300

TTCAACATTG AGGAGAACTG GAAAGGGCAC CACGAGACTG ACCACACAGA CCTTAACGGC 360

TGGATCTGGG AGCGCAAGTA TGAGGTGGAT TCGCTTTGCT ATCCTTTGCA GTTGGCTTAT 420

CTCCTCTGGA AAGAGACTGG CGAGACTAGT CAGTTTGATG AGATTTTTGT CGCAGCGACT 480

AAGGAAATTC TCCATCTGTG GACGGTGGAA CAAGACCACA AGAACTCTCC TTATCGTTTT 540

GTCCGAGATA CGGACCGTAA GGAAGACACC TTGGTAAATG ATGGCTTTGG ACCTGACTTT 600

GCAGTGACAG GTATGACTTG GTCAGCTTTT CGTCCGAGTG ATGACTGTTG CCAGTATAGT 660

TACTTGATTC CGTCAAATAT GTTTGCTGTA GTAGTCTTGG GTTATGTGCA AGAAATCTTC 720

GCAGCATTAA ACCTAGCTGA TAGCCAGAGT GTTATTGCTG ATGCCAAGCG TCTTCAGGAT 780

GAAATCCAAG AAGGAATCAA AAACTACGCT TACACCACCA ACAGCAAGGG CGAAAAGATT 840

TACGCTTTTG AAGTGGATGG CCTAGGAAAT GCCAGCATCA TGGATGATCC AAATGTACCA 900

AGTCTACTAG CTGCGCCCTA TCTGGGCTAC TGTTCGGTCG ATGATGAAGT GTATCAAGCT 960

ACTCGTCGTA CCATTTTGAG CTCTGAAAAT CCATACTTCT ACCAAGGAGA ATACGCAAGC 1020

GGTCTCGGCA GTTCTCATAC CTTCTATCGC TATATCTGGC CAATCGCCCT TTCTATCCAA 1080

GGCTTGACAA CAAGAGATAA GGCAGAGAAA AAATTCTTGC TGGATCAGCT GGTTGCCTGC 1140

GATGGTGGTA CAGGTGTCAT GCACGAAAGC TTTCATGTAG ATGATCCGAC CCTCTACTCT 1200

CGTGAATGGT TCTCCTGGGC TAACATGATG TTCTGTGAGT TGGTCTTGGA TTACTTGGAT 1260

ATTCGCTAAG GGGCTCGCTT TAGCTCAACC GATTCTTATC AGAATCACAA GTTTACATTT 1320

AAAACGTTAA AATTTAAATT TAGAATGAGG TTTTACTTCA TGGAAAATGT TGTTGTACAT 1380

ATTATCTCAC ATAGTCACTG GGATCGTGAG TGGTACTTGC CTTTTGAAAG CCATCGTATG 1440

CAGTTGGTGG AATTGTTTGA CAATCTCTTT GATCTCTTTG AAAATGACCC TGAGTTCAAG 1500

AGTTTCCACT TGGATGGACA AACTATTGTC CTTGATGACT ACTTACAAAT TCGCCCTGAA 1560

AATCGCGACA AGGTCCAACG CTACATTGAC GAGGGCAAAC TTAAAATTGG TCCCTTTTAC 1620

ATCTTGCAGG ATGACTACTT GATCTCCAGT GAAGCCAATG TCCGCAATAC CTTGATTGGT 1680

CAACAAGAAG CTGCCAAATG GGGTAAATCA ACCCAGATTG GCTACTTTCC AGATACCTTT 1740

GGAAATATGG GACAAGCGCC TCAAATTCTT CAAAAATCAG GCATTCACGT GGCGGCCTTT 1800

GGTCGTGGTG TGAAGCCGAT TGGATTTGAC AACCAAGTCC TTGAAGATGA GCAGTTTACG 1860

TCTCAGTTTT CAGAAATGTA CTGGCAGGGT GTGGATGGTA GTCGTGTTTT AGGTATTCTC 1920

TTTGCCAACT GGTACAGTAA CGGGAATGAA ATTCCAGTTG ACAAAGATGA GGCCTTGACC 1980

TTCTGGAAAC AAAAATTGTC AGATGTGCGT GCCTACGCTT CGACCAACCA ATGGTTGATG 2040

ATGAACGGCT GTGACCACCA GCCTGTACAG AAAAATCTGA GCGAAGCCAT TCGTGTGGCA 2100

AATGAACTCT TCCCGGATGT AATCTTTGTT CATAGTTCTT TTGATGAATA TGTTCAAGCT 2160

GTAGAAGGTG CGCTTCCTGA ACACTTATCA ACTGTTACAG GCGAGTTGAC CAGTCAGGAA 2220

ACAGATGGCT GGTACACACT TGCCAACACT TCTTCATCCC GCATTTACCT AAAACAAGCC 2280

TTCCAAGAAA ATAGCAACCT CCTAGAGCAA GTGGTAGAAC CCTTGACTAT TATCACTGGT 2340

GGACACAACC ACAAGGACCA GTTGACCTAT GCTTGGAAAA CACTTTTGCA GAATGCGCCA 2400

CATGATAGTA TCTGTGGCTG TAGCGTGGAC GAAGTTCACC GCGAGATGGA AACGCGTTTT 2460

GCCAAGGTCA ACCAAGTAGG AAACTTTGTT AAAAGTAACT TGCTCAACGA GTGGAAGGGT 2520

AAAATTGCTA CGGATAAGGC TCAAAGTGAC TATCTCTTTA CTGTCATTAA CACAGGCTTG 2580

CATGATAAGG TCGATACTGT CAGCACAGTG ATTGATGTGG CGACTTGTGA TTTCAAGGAA 2640

TTGCACCCAA CAGAAGGCTA CAAAAAGATG GCTGCTCTTA TCTTGCCAAG TTACCGTGTG 2700

GAGGACTTGG ATGGTCGTCC TGTAGAGGCT ACAATCGAAG ACCTCGGAGC TAATTTTGAG 2760

TATAATTTAC CAAAAGACAA GTTCCGCCAA GCTCGTATTG CTCGTCAAGT GCGCGTGACC 2820

ATTCCAGTTC ACCTAGCGCC GCTTTCTTGG ACAACCTTCC AATTGCTGGA AGGAAAACAA 2880

GAACACCGTG AGGGTATTTA CCAAAACGGA GTGATTGATA CACCATTCGT AACGGTGAGT 2940

GTGGATGACA ACATCACAGT CTATGACAAG ACAACTCACG AAGCCTATGA AGACTTTATC 3000

CGCTTTGAAG ACCGTGGGGA CATCGGAAAC GAGTATATCT ATTTCCAACC AAAAGGAACA 3060

GAGCCAATCT TTGCAGAGCT TAAGGGCCAC GAGGTCTTGG AAAACACAGC TTGCTATGCT 3120

AAAATCTTGC TCAAACATGA ATTGACCGTG CCTGTCAGTG CGGATGAAAA GCTAGAAGAA 3180

GAGCAACAAG GTATCATCGA GTTTATGAAG CGTGAGGCTG GACGGTCAGA AGAATTGACA 3240

AACATTCCTC TGGAAACTGA GTTGACTGTC TTCGTTGACA ATCCACAAAT CCGCTTCAAG 3300

ACTCGCTTTA CTAACACTGC CAAGGATCAC CGTATCCGTC TCTTGGTCAA GACTCATAAC 3360

ACGCGTCCAA GCAATGATTC TGAAAGTATC TATGAGGTGG TGACACGACC AAACAAACCA 3420

GCTGCTTCAT GGGAAAACCC TGAAAATCCT CAACACCAAC AAGCTTTTGT CAGTCTGTAT 3480

GACGATGAAA AAGGGGTGAC TGTATCCAAC AAGGGATTGA ATGAATACGA AATCCTTGGG 3540

GATAACACCA TTGCCGTGAC CATTTTGCGT GCATCAGGTG AGCTAGGTGA CTGGGGCTAC 3600

TTCCCAACGC CAGAAGCACA ATGCTTGCGG GAGTTTGAAG TCGAGTTTGC ACTTGAATGC 3660

CACCAAGCCC AAGAACGCTT CTCAGCCTAT CGTCGTGCCA AAGCCTTGCA GACACCGTTT 3720

ACCAGCCTTC AGCTTGCTAG ACAGGAAGGA AGCGTGGTTG CGACTGGTAG CCTCTTGAGC 3780

CATTCTGTTC TCAGCATACC GCAAGTTTGT CCAACAGCCT TTAAGGTAGC TGAAAATGAA 3840

GAAGGCTATG TGCTTCGTTA CTACAATATG TGTAGTGAAA ATGTACGTGT GCCAGAAAGT 3900

CAACATCTCT TCCTTGACCT ACTTGAACGA CCATACCCAG TTCATTCAGG ACTATTGGCT 3960

CCACAAGAGA TTCGTACAGA ATTCATCAAA AAAGAAGAAA TTTAATTTCA AAAAGTAAAC 4020

ATCAAAAGAA AGGAGGGGCG AAAAAGTAAG AACTAACTGC TGATTCGCCC CTTTTATGGT 4080

AAAAACAATG ACCATTGCAA CGATTGATAT CGGAGGGACT GGGATTAAGT TTGCCAGTCT 4140

GACTCCTGAT GGGAAAATAC TGGATAAGAC AAGTATTTCA ACGCCTGAAA ACTTGGAGGA 4200

TTTACTAGCG TGGCTAGATC AACGCTTGTC AGAACAGGAT TACAGTGGGA TTGCTATGAG 4260

CGTTCCAGGT GCAGTCAATC AAGAGACAGG TGTGATTGAT GGCTTCAGTG CGGTGCCCTA 4320

CATCCATGGC TTTTCTTGGT ATGAGGCGCT TAGCTCTTAT CAGCTACCTG TCCATTTAGA 4380

AAATGATGCC AACTGCGTTG GACTCAGTGA ACTACTAGCT CATCCAGAGC TTGAAAATGC 4440

AGCCTGTGTC GTGATTGGGA CAGGGATTGG CGGAGCCATG ATTATCAATG GTAGACTTCA 4500

TCGAGGTCGC CACGGTCTGG GTGGAGAATT TGGCTACATG ACAACCCTTG CCCCTGCTGA 4560

AAAACTTAAT AACTGGTCGC AACTAGCATC AACTGGGAAT ATGGTACGAT ACGTGATTGA 4620

AAAATCTGGT CATACTGATT GGGACGGTCG CAAGATTTAC CAAGAGGCCG CAGCTGGTAA 4680

TATCCTTTGT CAAGAAGCCA TTGAGCGCAT GAACCGCAAT CTGGCGCAAG GCTTGCTCAA 4740

TATCCAGTAT CTGATCGATC CAGGTGTCAT CAGTCTGGGT GGCTCTATCA GTCAAAATCC 4800

AGATTTTATC CAAGGTGTCA AGAAGGCTGT TGAAGACTTT GTCGATGCCT ACGAAGAATA 4860

CACGGTCGCA CCAGTTATCC AGGCCTGCAC CTATCACGCA GATGCCAATC TCTACGGTGC 4920

TCTTGTCAAC TGGCTACAGG AGGAAAAGCA ATGGTAAGAT TTACAGGACT TAGTCTCAAA 4980

CAAACGCAAG CTATTGAGGT TTTAAAAGGT CACATTTCTC TACCAGATGT GGAAGTGGCT 5040

GTCACTCAGT CTGACCAAGC ATCTATCTCT ATCGAGGGTG AGGAAGGTCA CTATCAATTG 5100

ACCTACCGCA AACCTCACCA ACTTTATCGT GCCTTGTCCT TGTTGGTAAC AGTTCTAGCA 5160

GAAGCTGATA AAGTAGAGAT TGAGGAACAA GCAGCTTACG AAGATTTGGC TTACATGGTT 5220

GACTGTTCTC GAAATGCGGT GCTGAATGTG GCTTCTGCCA AGCAGATGAT TGAGATATTG 5280

GCTCTCATGG GCTACTCAAC CTTTGAGCTT TACATGGAAG ACACTTACCA GATTGAAGGG 5340

CAGCCTTACT TTGGCTATTT CCGTGGAGCT TATTCAGCAG AGGAGTTGCA GGAAATCGAA 5400

GCCTATGCCC AACAGTTTGA CGTGACCTTT GTACCATGCA TCCAGACCTT GGCCCACTTG 5460

TCGGCCTTTG TCAAATGGGG TGTCAAGGAA GTGCAGGAGC TCCGTGATGT AGAGGACATT 5520

CTTCTCATTG GCGAAGAAAA GGTTTATGAC TTGATTGATG GCATGTTTGC CACGTTGTCT 5580

AAACTGAAGA CTCGCAAGGT CAATATCGGG ATGGACGAAG CCCACTTGGT TGGTTTGGGA 5640

CGCTACCTGA TTCTGAACGG TGTTGTGGAT CGTAGTCTCC TCATGTGCCA ACACTTGGAG 5700

CGCGTGCTGG ATATTGCTGA CAAATATGGT TTCCACTGCC AGATGTGGAG TGATATGTTC 5760

TTCAAACTCA TGTCAGCGGA TGGCCAGTAC GACCGTGATG TGGAAATTCC AGAGGAAACT 5820

CGTGTCTACC TAGACCGTCT CAAAGACCGT GTGACTCTGG TTTACTGGGA TTATTATCAG 5880

GATAGCGAGG AAAAATACAA CCGTAATTTC CGCAATCATC ACAAGATTAG CCATGACCTT 5940

GCATTTGCAG GGGGAGCTTG GAAGTGGATT GGCTTTACAC CTCACAACCA TTTTAGCCGT 6000

CTAGTGGCTA TCGAGGCTAA TAAAGCCTGC CGTGCCAATC AGATTAAAGA AGTCATCGTA 6060

ACGGGTTGGG GAGACAATGG TGGTGAAACT GCCCAGTTCT CTATCCTACC AAGCTTGCAA 6120

ATCTGGGCAG AACTCAGCTA TCGCAATGAC CTAGATGGTT TGTCTGCGCA CTTCAAGACC 6180

AATACTGGTC TAACGGTTGA GGATTTTATG CAGATTGACC TTGCCAACCT CTTACCAGAC 6240

CTACCAGGCA ATCTCAGCGG TATCAATCCC AACCGCTATG TTTTTTATCA GGATATTCTT 6300

TGTCCGATTC TTGATCAACA CATGACACCT GAACAGGACA AACCGCACTT CGCTCAGGCT 6360

GCTGAGACGC TTGCTAACAT TAAAGAAAAA GCTGGAAACT ATGCCTATCT CTTTGAAACT 6420

CAGGCCCAGT TGAATGCTAT TTTAAGTAGC AAAGTAGATG TGGGACGACG CATTCGTCAG 6480

GCCTACCAAG CGGATGATAA AGAAAGTTTA CAACAAATCG CCAGACAAGA ATTACCAGAA 6540

CTTAGAAGCC AAATTGAAGA CTTCCATGCC CTCTTTAGCC ACCAATGGCT GAAAGAAAAC 6600

AAGGTCTTTG GTTTGGATAC AGTTGACATC CGTATGGGCG GACTCTTGCA ACGCATCAAA 6660

CGAGCAGAAA GCCGTATCGA GGTTTATCTG GCTGGTCAGC TTGACCGCAT CGACGAGCTG 6720

GAAGTTGAAA TCCTACCATT TACTGACTTC TACGCAGACA AGGATTTCGC AGCAACTACA 6780

GCCAACCAGT GGCATACCAT TGCGACAGCG TCGACGATTT ATACGACTTA ATATTCTTCG 6840

AAAATCTCTT CAAACCACGT CAGCTTCCAT CTGCAACCTC AAAACAGTGT TTTGAGCAAC 6900

CTGCAGCTAG CTTCCTAGTT TGCTCTTTGA TTTTCATTGA GTATAAAAAC AAGAACACCT 6960

TGCTTGGCGC AGGGTGTTTC GCGTGAAACA GAAGAATTAT CTGGTTTCAA ATGCTACAGT 7020

TAGACAAACT TATGATAAAA TAGCAGAAAG TGAATGTTTC CTAAGAGCAA TTGGAGGTAT 7080

TATGCTACAC TTAAAATTAG TAAAACAAGA AATAGAAGCT GAAAAGCCAG CATCTGTAGA 7140

AGCTTGGATC ATTTCCGTCA AATTTAAAAA AGGTTGCTAC CGACATATAT AGATTCCAAA 7200

AACAAAAACG TTAGCGGAAC TAGCAGATGT GATTTTATGG AGTTTTGATT TTGCAAATGA 7260

TCATGCTCAC GCATTTTTCA TGGATAATGT TGAGTGGAGT CATGCAGATT CTTACTTTCG 7320

TAGCTTTGTT AGTGACGATG TTGAAGAACG TTACACAGAA AATGTCTATC TGGATAGCCT 7380

AAGTGTCAAA CAAAAATTTA AGTTTATTTT CGACTTCGGT GATGATTGGC GTTTTGAATG 7440

CCAAGTGCTG AGAGAAATCG AGACAGAGGA CGAAGAAGCT TATCTCGTAC GTTCGGTTGG 7500

AACGTCGCCA GAACAATATC CAGATTATGA TGGTTTTGAC TATGAAGAAT GGTAAAATTG 7560

AAATCAGTCT GTGTAGGCTT AGTATTTCAA TAGACTTCCT GCAAAACTAG AATCCTAGTT 7620

CATGATTGAT AATACCAGCA ATCAAATTCA TTCGTAATCC GAAGCGTTTA CGATGATTTC 7680

GATAGGTTGT TGAAAACATT TTAAACGTTT TTACTTTGGC AAAGATGTTC TCAACCTTGC 7740

TTCTCTCCTT AGATAGCGCA TGGTTATAGG CTTTATCTTC AGCTGTTAGT GGCTTGAGTT 7800

TGCTGGATTT ACGTGAAGTT TGTGCTTGAG GACATATCTT CATGAGCCCT TGATAACCAC 7860

TGTCAGCCAA GATTTTACCA GCTTGTCCGA TATTTCTGCA ACTCATTTTG AACAACTTCA 7920

TATCATGACA ATAGTTCACA GTGATATCCA AAGAAACAAT TCTCCCTTGA CTTGTGACAA 7980

TCGCTTGAGC CTTCATAGCG TGAAATTTCT TTTTACCAGA ATCATTCGCT AATTCTTTTT 8040

TTAGGGCGAT TGATTTTTAC TTCCGTCGCA TCAATCATTA CCGTGTCCTC AGAACTAAGA 8100

GGAGTTCTTG AAATCGTAAC ACCACTTTGA ACAAGAGTTA CTTCAACCCA TTGGCTCCGA 8160

CGGATTAAGT TGCTTTCGTG AATACCAAAA TCAGCCGCAA TTTCTTCATA AGTGCGGTAT 8220

TCTAGGCTTA ATTTAGGTTT TCGTCCACCT TTTGCGTGTT TAAGTTGATA AGCTGTTTTT 8280

AATACAGCTA ACATCTCTTT AAAAGTCGTG CGCTGAACAC CAACAAGACG CTTAAATCGT 8340

GTATCAGTTA ATTGTTTACT TGCTTCATAA TTTCGCAGGG AGTCTATTGA CTCTTTGGTA 8400

GGTGTCAATG TTTTTTTCAT CTATCCCGAG AATTATTTTC CCGCCATTTG TATTTGCAAA 8460

TGCTGAGTAG GTTTCCCAGA AAGACTCTGG AAGATTGTTT TTAGCTTTTT TGTATTCTAA 8520

ATCAACCCCT TCAAATTTTA AGTCCATATT TTTCCTTTAC ATCTGTTTTT TGTGGTTCTG 8580

GTATTTGTTC AAGTTGAGTG ATAATATAGC GAATTGAATT TCGAGAGTTT TTACTCAGTT 8640

AATTTCTTTT TTAACCC 8657

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11384 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

TCTATTTTGG GTATAGACTT ACCTATAAAG AAAAATATCT ATACACTGCC TTACTAGCTA 60

TACTGAACGA GTCAACAAAA ACGATATATA TTGATGATAT AAATACAGCA AGATTTTTTA 120

ACTTCTTTGG CAATGATATT CCTAATTCGT CTTTAAAAAA AATTGACTAT ATCGCACCTT 180

CAGAAATTGT TTCATTTAGT ACGTACGTTC GACAACGTTC TAAAGTAATT CCTAAAATTT 240

TGGAACATAT ATTAAAATCA AGTTTTTTAT TAGAGAATAT AGATGTTTCT GGTTACACTG 300

TAAATATTTT AGAAGATCAA TTAACAAAAC ATAGAACAAT CAAAATTAGT AAAAACTAAC 360

TGGTTGATCT CATGTATAAA TACCTAACAA AACCACGCGC CTTGCCTGCT GATGGAAAGA 420

AAGGTACAAA TACATGAATA TCAAAGAAAA AATCAAAAAG AATGGCCAAA GAGTTTATTA 480

TGCTAGTGTT TATCTAGGCG TTGACCAACT AACGGGCAAA AAAGCCCGTA CAACTGTTAC 540

AGCAACCACT AAAAAGGGCG TTAAAGTAAA AGCGCGTGAT GCGATCAATA CTTTTGCTGC 600

TAATGGCTAT ACAGTTAAAG ACAAGCCGAC AATTACAACA TATAATGAGC TTGTAAAAGT 660

TTGGTGGGAT AGTTACAAGA ATACAGTTAA GCCAAATACT CGCCAATCCA TGGAGGGATT 720

GGTTAGAGTG CATTTATTGC CTGTATTTGG CGATTACAAG CTATCTAAAC TTACTACGCC 780

TATTCTTCAA CAGCAAGTAA ACAAATGGGC TGACAAGGCA AATAAAGGCG AAAAAGGGGC 840

ATTTGCTAAC TACTCTTTGC TCCATAACAT GAATAAGCGT ATTTTGAAAT ATGGCGTAGC 900

TATCCAGGTA ATACAATACA ACCCAGCTAA TGATGTCATC GTTCCACGCA AACAGCAAAA 960

AGAAAAGGCT GCTGTCAAAT ACTTAGACAA CAAAGAATTA AAACAGTTTC TTGATTATTT 1020

AGATGCTCTG GATCAATCAA ATTATGAGAA CTTATTTGAT GTTGTTCTGT ATAAGACTTT 1080

ATTGGCCACT GGTTGCCGTA TTAGTGAGGC TCTGGCTCTT GAATGGTCTG ATATTGACCT 1140

AGAAAGCGGT GTTATCAGCA TCAATAAGAC ACTAAACCGC TATCAGGAAA TAAACTCACC 1200

TAAATCAAGC GCTGGTTATC GTGATATACC AATAGACAAA GCCACATTAC TTTTACTGAA 1260

ACAATACAAA AACCGTCAAC AAATTCAGTC TTGGAAATTA GGCCGATCTG AAACAGTTGT 1320

ATTCTCTGTA TTTACGGAGA AATATGCTTA TGCTTGTAAC TTACGCAAAC GCCTAAATAA 1380

GCATTTTGAT GCTGCTGGAG TAACTAACGT ATCATTTCAT GGTTTCCGCC ATACACATAC 1440

TACTATGATG CTCTATGCTC AGGTTAGCCC GAAAGATGTT CAGTATAGAT TAGGCCACTC 1500

TAATTTAATG ATCACTGAAA ATACTTACTG GCATACTAAC CAAGAGAATG CAAAAAAAGC 1560

CGTCTCAAAT TATGAAACAG CTATCAACAA TTTATAAAAA ATAAGGGTGA CCCATTTCCG 1620

GGCTACCCTC TTACTATACC AAAAATTAGT AGGGGTAGTA AAAAGGGTAT TAAATTATAA 1680

AAAGCACTAA GGGAAAGCGC CCCAAAGTGC TTATTTCAAA GGCTTTATAG CCTATAATCA 1740

CATAAAGAGA TTATTTTTTA AGGTTGTAGA ATGATTTCAA TCCACGATAT TCAGCTACTT 1800

CACCAAGTTG GTCTTCGATA CGAAGCAATT GGTTGTATTT AGCGATGCGG TCTGTACGTG 1860

AAAGTGAACC AGTCTTGATT TGTCCTGCGT TAGTTGCAAC TGCAATATCA GCGATTGTTG 1920

AATCTTCAGT TTCACCTGAA CGGTGTGATA CAACAGCAGT GTAACCAGCT TCTTTAGCCA 1980

TTTCGATAGC TTCAAAAGTT TCAGTAAGAG TACCGATTTG GTTAACTTTG ATAAGGATTG 2040

AGTTAGCAGC ACCTTCTTGG ATACCACGTG CAAGGTAGTC AGTGTTTGTT ACGAAGAAGT 2100

CGTCACCAAC AAGTTGTACT TTCTTACCAA GACGTTCAGT AAGAGCTTTC CAACCATCCC 2160

AGTCGTTTTC ATCCATACCA TCTTCAATAG TGATGATTGG GTATTTGTTA ACCAATTCTT 2220

CAAGGTAGTC GATTTGTTCT GCAGATGTAC GAACAGCAGC ACCTTCACCT TCAAATTTAG 2280

TGTAGTCGTA AACTTTACGT TCTTTATCGT AGAATTCTGA TGAAGCACAG TCAAATCCGA 2340

TAAATACGTC TTTACCTGGT ACATATCCAG CAGCTTCAAT CGCAGCAAGG ATAGTTTCAA 2400

CACCATCTTC AGTTCCTTCG AAACGAGGAG CGAATCCACC TTCGTCACCT ACGGCAGTTT 2460

CCAAACCACG TGATTTAAGG ATTTTCTTAA GAGCGTGGAA GATTTCAGCA CCGTAACGAA 2520

GGGCTTCTTT AAATGTTGGC GCACCAACTG GCAAGATCAT GAACTCTTGG AAAGCGATTG 2580

GAGCGTCAGA GTGAGAACCA CCGTTGATGA TGTTCATCAT TGGAGTTGGA AGAACTTTAG 2640

TGTTGAATCC ACCAAGATAG CTGTAAAGTG GGATTTCAAG GTAGTCAGCA GCAGCACGAG 2700

CTACAGCGAT AGACACACCG AGGATTGCAT TCGCACCCAA TTTACCTTTG TTAGGAGTAC 2760

CGTCAAGTGC GATCATAGCA CGGTCAATAG CTTGTTGATC ACGTACATCG TAGCCAATGA 2820

TAGCTTCAGC AATGATGTTG TTTACGTTGT CAACAGCTTT TTGTGTACCA AGACCACCGT 2880

AACGAGATTT GTCACCGTCG CGAAGTTCAA CTGCTTCGTG TTCACCAGTA GAAGCTCCTG 2940

ATGGAACCAT ACCACGTCCG AAAGCACCTG ATTCAGTGTA AACTTCTACT TCAAGTGTTG 3000

GGTTACCGCG TGAGTCTAGG ACTTCGCGAG CGTAAACATC AGTAATAATT GACATTTTTT 3060

ACTCTCCTTA TGAGTTAAAT TTTTTACACC TCTATAATAC CTTAAAACCC CTCCTTTTTC 3120

AAGAAAAAAC GTTATCTTTG TGCAACTTTT CCTTAACTTT ATAAAGTAAT CGCTTTCTTT 3180

TGTCTGTTTT ATTCTAACTT TTATGATATA CTGTTTTCAT GACAGATTTA TCAAAACAAT 3240

TACTTGAAAA AGCTCATGGT GGGTTAAAAA TAAATCCGGA TGAGCAAAGA CGCTATCTTG 3300

GTACTTTTGA GGAAAGAGTT CTTGGATATG TAGATATTGA CACAGCAAAT AGCCCTCAGT 3360

TAGAAAAAGG CTTTTTATTT ATTTTAGAAA ACCTTCAGGA AAAAGCAGAG CCACTATTTG 3420

TGAAGATTTC ACCAACTATC GAATTTGATA AGCAAGTTTT CTACTTAAAA GAAGCAAAAG 3480

AAACTGATAG TCAAGCCACC ATAGTATCTG AAGAGCATAT TACTTCTCCT TTTGGCCTGG 3540

TTATTCATAG CAATGCACCA GTTCAAGTAG AAGAAAAAGA CCTTCGACTT GCTTTTCCAA 3600

AACTTTGGGA AGTTAAAAAG GAAGAACCAG CCAAAACATC CTTATGGAAG AAATGGTTTA 3660

GCTAAATCTT GCACATATTT AATAAGTGCC CAATATTGGC AGCCGTGCGC TCCAGATAGA 3720

AACTGGCATT TTTCAAACTA TCTTCTAAAG GTTCACTTTT CTCCAAAATA GAAAAGACAG 3780

CTTGGATATT TTCAAATGGT AGGGGAGGTA AATCTTCAGC AAGACTACCG CAAATAGCAA 3840

TAACAGGAAC TCCAACAGGG GTTCTTTTTG CAACACCTAT AGGCGCTTTC CCAGCAAAGC 3900

TTTGACTATC AAGTCTTCCT TCTCCAACAA CAACCAAGTC AGCATCTGAA ACTTTCTTAT 3960

CAAAGTTGAT TAAGTCCAAG CAGGTATCAA TTCCAGACAC GATACTTGCC TGAGCAAAGG 4020

CACACAAACC ACCAGCAAGG CCTCCACCTG CTCCTGCTCC TTTAATTTCT AATGTTGCAG 4080

GTGAGAATTT TTCATAAAAA TCTTGGATCG CCTGATCTAC GACTGCAAAC ATAGTCGGAT 4140

GTAGACCTTT TTGATTGCCA AAAGTGTAAG TCGCACCTTG ATGACCACAT AAGGGACTCA 4200

CGACATCTGC TAAAATATGA ATTTGAACAC CTTCAGGAAT TTTATAGCAA TTTTCTGTTG 4260

AAACAGAAGC TAAGTTTAAT AAGGATTGAC CGGAAGCAGG CAAGACATTT CCATCCCTAT 4320

CATAAAATTG ATAACCTAAA CCAGCAGCAA TCCCCAGTCC TCCATCATTA CTGGCCGTGC 4380

CACCAACACC GATATAAATA TCTTTAATCC CTTTAGAGAT GAGATGAAGA ATCAACTCTC 4440

CAATACCACA AGTTTGGATT TGAAGTGGAT TTCGTTTCTC TAGCGGAATT TTTCCAAGAC 4500

CAACCAAGTC AGCTACTTCA AATAGTGCCA GTTCCCCTTT TTGAAAATAG CGCATGGCTT 4560

CTTTTTGTCC AAAAGGGTCT GTCACTTGGA TCCATTTTTC TTTTAGGTCA AGAGAATGTC 4620

GGATAGCATC TACAGTACCT TCTCCCCCAT CACCAACAGG GCAGAGGAGA CATTCTACAT 4680

CTGCTATCGA TTGTTGGAAG CCTCTTTTTA TTGCTTCAGC TACCTGTTGA GCTGTCAAGC 4740

TTTCCTTAAA CGAATCCGGT GCAATTACAA TCTTCATATT TTCCCTCATT CTAAACAGTC 4800

AATCAAAGGG AGAACTTCTA AAAAATCCCT CTTGTCAACA TGATGTGGTA TTTCTTTTTT 4860

GAGCACTTCT TTGGCACAAA AGGCGATTCC TAACTTCGCC GACTTCAACA TTAATAGATT 4920

ATTAACCCCA TCACCGATTG CCACCGTTCT TTCTTTAGAA AGTTTTAGTT TCTTTCTCCA 4980

TTTTTCCAGA GTCTCTTTTT TGACCTGGGG ACTTATAATT TGTCCAACTA ATTTTCCTGT 5040

TAAAAGACCT TCTTTGACTT CAAGCTAGTT GGCAGTGAAA TAGGCAATAC CAAGGGATTT 5100

TGCTAATCTC TCCAACTATT GGTGTAAATC CACCAGACAC CAGACCAACT AGGATGCCAT 5160

TCTTTTGGAG AATAGAGATG AACTCTGGGA CATTTAGCGA TAGATGAATT GAGTTGAAGA 5220

CGTTATCAAA GACCAAAATA GGAAGACCTT CCAACAAGGA CACTCTTTTT CTTAAACTGC 5280

TTTCAAAGAC CAACTCTCCT CGCATTGCTC GACTTGTAAT CTGCGAAATT TCCGCCTCAT 5340

GACCTGCCTC TCTCCCTAAA AGATCAATCA CTTCTTCTAG GATTAAGGTT CCATCTACAT 5400

CCAAAACACA CAAGCCTTTT ACTTGAGACA TCAGTTCTCC TCTCTAAACA GCCTAAAAAT 5460

CGTATGAAGT CATCATACGA TTTTATCTAT TAATTAACTA AACTATGGTA CAAGTCAAGG 5520

TATGACTTGC AGGCTGTATC CCATGAGAAG TCACTCTCCA TAGCTTGTTT TTGTAGGTTT 5580

CTCCAAATGT CTGGATGGTT TCTATACAAG TCCAATGCTG TTTGGAAAGT CCAATTTAAC 5640

CAATAAGGAG ATAGATTGTC AAAGCTAAAG CCAGTACCGC TTCCTTCGAT TGGATTGAAA 5700

GCGCGAACTG TATCTCGCAA GCCTCCAACT TCATGGACCA ATGGCAAGGT TCCATAACGC 5760

ATAGCCATCA TTTGAGACAA GCCACACGGT TCAAAACGAC TTGGCATGAG GAAGAGGTCA 5820

CAAGCAGCGT AGATTTCCTG AGCAAGTTTG ACATCAAAAG TGATATTTGT TGATAGCTTG 5880

TCTGGGTAAA TCTGAGCAAA CCATGAGAAA GCTCCTTCAA AGGCTGGATC GCCAGTTCCC 5940

AAAAGAACAA TCTGAACATC TTCTTGCAAG ATATGGTGAA GACTTTCGAC CACCACATCA 6000

AAACCTTTTT GACGTGTCAA ACGAGAAACA ATTCCCACCA GTGGAACGTC TGCTCTAACA 6060

GGCAAGCCAA CTCTTTCTTG CAATTTTGCC TTATTTTTGG CTTTCCCAGA CAAATCTTCC 6120

TGATTGAAAT GATAGTCTAA AAGAGCATCC GTCTGAGGAT TATAAAGATC AGCATCAATC 6180

CCATTCACGA TACCAGATAC TTTACCAGAC TCCATTTTAA GAATCTGATC CAAATTACAT 6240

CCAAACTGAC TAGTCATAAT TTCATGAGCA TAGCTAGGTG AAACGGTTGA AACACGGTTC 6300

GCATAGAGAA TACCTGCCTT CATCCAGTTC AGACAGTTGT TCCATCGAAG GGTGCCATCA 6360

GCGTAACGTT CAAAGCCAAC TCCAAACAAA TCACCCAACA TTCCTTCTGA AAATTGTCCT 6420

TGGAATTCTA AATTATGAAT GGTTAAAACT GTTTCAATGT CCTCATAGGC TTGAATCCAA 6480

CGGTATTTTT CCTTCAACAA GAAAGGAATC ATAGCTGTAT GGTAGTCATG AACATGGAGA 6540

AGATCAGGAA TAAAGTCAAT CCTTTCCATA GCCTCAATGG CAGCCAGTTG GAAAAAGGCA 6600

AAGCGTTCTC CGTCATCAAA ATCACCGTAA ACATGACCAC GGAAGAAATA ATATTGATTG 6660

TCAATAAAGT AGAAGGTTAC ACCATTTAAT ACTGTTTTCT TAATTCCACA ATACTGTCTG 6720

CGCCAACCAA CGCTCACCTC AAAATGAAGC ACATCTTCAA TCTGATTTCC AAATTTAGCC 6780

TCTACCATAT CATAGTAGGG TAAAATCACT GCAACTTCGT GCCCAGCTTT TACCAGTGAT 6840

TTTGGAAGAG CGCCAATGAC GTCTCCCAAA CCACCTGTTT TTGAAAAGGG TGCACCCTCT 6900

GCTGCTACAA ATAAAATTTT CATGAATGAA TATCCTCTGT TACTTTAGCA CCTTTCTTAA 6960

CCACAACTGG ATGTTCTGCA GTTCCTCGAA TCACAACACC ATGCTCAACT TCAACCCCTT 7020

TGTCCAAGAT AGCATATTCG ACCTGAGCCC CTTCTCCAAT AACAACACGA GGGAAGAGCA 7080

GGCTATCTTT AACCAAGCTA TCCTTATGGA CATGAATATT ACGTGATAGA ACAGAATTAG 7140

CTACTTGACC TTCAATAATA CTACCAGAGG CAAACTGAGA AGTGCTTACC TTAGATGTAT 7200

TAGCATAGTA AGTTGGCTCT TCGTTTTTGA CCTTTGTATA AATCTTTTGG TTTGGTGAGA 7260

AAAGAGAATA GAATTTTTGT GATTCAAGCA TATCGATATT CGCTTGATAA TAAGATTTAA 7320

CAGAGTGAAT ATTGGCTAGA TAGCCCGTGT ACTCGTAGGC GAAAGCTCCC TCTTTTACAG 7380

CCAAATCCCG TAAAACATAG CGCAATTTCT CTGGATGTTC TTTTTTAGCT TCTTCTTCCA 7440

AGTGTTCAAT CAACCAAGGT GTATCAACGA CAAAGATATC TGTAGACATA TTGAACGTTT 7500

CAGCTGTTGA CTTGCTATCA AAGAGTTTAT GAGAAAGAAC ATGGTCTGTT TCATCTACAT 7560

CCAAGATTGC ATTTACTTCT GAAATATCTT TCTTAGCTAG TTTTTTATAA ACTACAGTGA 7620

TAGGCTCTTT TGTTGTACTA TGTAGGTGGA AAACTTGGTT CAAATCAATG TTAATAAGAA 7680

CATCGCAGTT GAGGGCAACC GTTTGGTTTG AGCCAGAACG TTTCAAATAA GTAAGAAGCT 7740

GTTGGTAGTA TTCTTTTCCA ACTGTACTAC TTTCTACACG GGTATTGTAA ATTCCTAGAT 7800

AGTAATGGCT AAGAAGGGTT GATAAGCCCC ACTCGCGTCC TGAACGAATA TGGTCAAATA 7860

CTGAGCTGAT ATTATCCTGC TGGAAAATAC CAAAGACACT ACGAACACCT GCATTAGCAA 7920

GGCTTGAAAG TGGGAAGTCA ATCAAACGAT ATTTCCCACC AAATGGCAAA CTTGCTACTG 7980

GACGGTGGTC CGTCAATGTC GACATATTGT GAAAACCAAC TGTATTTCCT AAAATGGCAG 8040

AATATTTATC AATCTTCATC TGTTGCTACC CCCACTACTT CATTATATCC TACAACTTGT 8100

ACTTCATCTG TTCCATCAAT TTCGACACCG TCAGAAATAA TCGCACCTTC ACCAATAATG 8160

GCACGTTTAA TCTTAGCTCC TTGACCAATG ATAGCTCCAC TCATGATAAC TGAATCAAGG 8220

ACTTCCGCTC CTTCGCGAAC TTGCGCGCCT GTTGAAAGGA TAGAATGTTT AACAGTTCCA 8280

TCAACGAAAC ATCCGTCTAC AACTAATGAG TCTTCCACAT GAGCATTTGC CCCGAGGAAG 8340

TTTGGTGGTG AAATCAAGTT TCTTGAGTAA ATCTTCCATT GACGGTTACG ACTATCCAAG 8400

GCATTTTCTG GAGAAATATA CTCCATGTTC GCTTCCCAAA GTGACTCAAT AGTACCAACA 8460

TCTTTCCAAT AACCACTAAA TTCGTAAGCA TAAACACTTT CACCTGACTC AAGGTAATTT 8520

GGAATGACAT TTTTACCAAA GTCTGACATG CCAACCTTGC TCTTTTCAGC AGCGACTAAC 8580

ATATTACGAA GGCGTTGCCA ATCAAAAATG TAGATTCCCA TAGAAGCTTT TGTAGATTTA 8640

GGTTGAGCTG GTTTTTCTTC AAATTCAACA ATACGATTGT TAGCATCTGT GTTCATGATA 8700

CCAAAACGGC TTGCTTCTTT AAGAGGGACG TCTAAAACTG CTACTGTCAA GCTGGCATTA 8760

TTATCCTTAT GAGACTGGAG CATATCATCA TAGTCCATTT TGTAGATGTG ATCCCCAGAC 8820

AAAATCAAGA CATACTCAGG ATTGACACTG TCGATATAGT CGATATTTTG GTAAATAGCG 8880

TGACTAGTCC CCTCAAACCA ACGATTTCCT TCACTTGCAG AATAAGGTTG AAGAATAGAG 8940

ACACCTGAAT TAATACCGTC TAGTCCCCAG CTTGAACCAT TCCCAATATG GTTGTTGAGA 9000

GCAAGTGGTT GATACTGTGT AACGACCCCA ACATTGTGAA TCCCTGAGTT GGCACAGTTT 9060

GATAGGGCAA AGTCAATGAT ACGGTAGCGC CCACCAAATT GCACAGCTGG TTTTGCGATG 9120

CTTTGAGTGA GTTTACCGAG ACGAGTTCCT TGCCCACCAG CAAGAATCAA AGCTAACATT 9180

TCATTTTTCA TTTTCTACTC CTTTTTGGTT TTTATTTGTG ACGGTTTTAG TAGATTTCAA 9240

GCGACGTTTG ATTTTCCATA CACTTGCTCC CATAGCCGGT AGGGTAAAGG TTAAGGTCTG 9300

CTCATAATCT TTCCATAGTC CTTCTTGCGT TTGAACAGTT TGATTATGTT CTTTCCAAAC 9360

GCCTCCCCAC TCTTCCAACT CAGTATTCCA TACTTCTTCG TAAATTCCTG CAACGGGTAG 9420

TCCGATTGTA AAATCTTTCC GCTCAACAGG TACCATATTA AAGATACAGA CTAACATTTC 9480

TCCCTTTTTA CCCTTACGAA TAAAGGAAAG AACACTCTGG TCTCGATTAT CCGCATCAAT 9540

GATTTCAATA CCATCATAGC TGGTATCAAT TTCCCACAGA CAGCGATGAT CTTTGTAAAA 9600

CTGGTTTAGC TGAGAAGCGA AATACTTCAT CTTAGCATTC ATTGGGTCTT CTAGGTTAGA 9660

CCATTCCAAC TGTTCTTCAG ATTTCCATTC TAGGAATTGA CCGTATTCGC TACCCATGAA 9720

GAGCAATTTC TTACCAGGGT GACAAATTTG GTACGTATAG AGATTGCGCA AGCCTGCGAA 9780

TTGATTGTAA CGATCTCCCC ACATCTTATG CATCATACTC TTCTTGCCAT GAACCACTTC 9840

ATCGTGCGAG AATGGCAAGA GATAATTCTC CTTGAAAACA TACATAAAGC TGAAAGTCAC 9900

CAGGTTAAAG TCATATTTAC GATAGATCGG ATCTTCTTCG TAGAAACGGA GGATATCATT 9960

CATCCAGCCC ATGTTCCATT TGTAGTCAAA TCCTAGACCA CCAATCTCTT TCATTCCCGT 10020

AATCTTGATC GCAGACGAAC TTTCTTCTGC AATCATCATC ACATCTGGAT ATTCTAACTT 10080

AATAACCTCA TTCAAGCGCT GAAGGAAATA ATAACCTTCA TAGTTGAGAT TTCCGCCATC 10140

TTTATTAGGT GTCCATGGAG CATCATCATA GTCCAAATAG AGCATGTTGC TAACAGCATC 10200

CACACGAATA CCATCCAAAT GATAGACATC AATCCAATGC TTAATGCAAG AAATTAAGAA 10260

GGACTGGACT TCATTTTTTC CAAGGTCAAA ATTAAGGGCA CCCCAACCAT GGTTATGAGC 10320

CTTATTATGG TCTTGGTATT CAAAAGTCGG TGTCCCATCA TAATAGGCTA AGGCATCATC 10380

GTTGATGGTA AAGTGACTGG TACCCAGTCC ACAATAACCC CAATATTATG GGTATGACAC 10440

TCCTCGACAA AATCTTGAAA CTCCTCTGGT CGGCCATAAG CATGCTCTAA AGCGAAGTAA 10500

CCCATAAGCT GATACCCCCA ACTCAAGCCC AAAGGATGGG ACATCAAGGG CATAAACTCA 10560

ATATGAGTAT AGTTCATTTC AACGAGATAA GGAATGAGTT CATCCTTGAG CTGGGCAAAA 10620

CTATAAGGAC TGCCATCAGA ATTTCTTTTC CATGATCCAG CGTGAACTTC ATAAATATTG 10680

ACAGGACGCT CTTCAAAGCC CCAACGTTTT CTTCGTGCCA GCCAAAGTCC ATCCTTCCAT 10740

TTCTTCTCAG GAAGCTCTGT TACGATTGCC CCTGTTCCTG GACGAGCCTC ATACCTGACA 10800

GCAAAAGGGT CAATCTTCAT CAGTTGATGA CCATTTTGAC GTGTGACATG ATATTTGTAA 10860

ATATGCCCTT CTTGAGCCAT ATTGGTAAAG ACTTCCCAGA CCCCAAAATC ATTTCTTACC 10920

ATTGGAATCT GATTTTCAAT CCAGTTGGTA AAATCACCAA CCAAGTGAAC AGCCTGAGCA 10980

TTAGGTGCCC AAACACGGAA GGTATAGCCA TGCTCTCCAT TTAGTTCTTC CCTATGTGCT 11040

CCTAGATAAT GTTGGAGATA AAAATTTTCA CCCGTCATAA AGGTTTTTAA TGCTTCTCTA 11100

TTATCCATAT ACTCCCCTTC TCCTGTAAGC GTTTTCTATG TTTTTATTAT ACTACCTTTT 11160

TAGAGAAGAT TCAAGTAAAT TACTATACTT CTTTAATTAT TTTGAAAATC TACAACAAGT 11220

TCACTTACTC GTTCAATTGT AAATCAATAT TTTTTCAAAA AATTGCGAAA ACGCCTTTCT 11280

TTTTCTACTA TAGTGAAATG AAATAAAACA TGCGCAAATC GATTAAGGAA TTTAATCTAA 11340

TTTCTAACAA TGTCTTAGAA ATCAAAGTGT ACTATTTTAA CTCC 11384

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7577 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

TGTTGATTTG TTACTAGACG TTGACCAACG TCCTTCGGCT GGAAAAGGAA TTCTCCTTAG 60

TTTCCAACAC GTTTTCGCCA TGTTTGGTGC GACCATCTTG GTACCATTGA TTTTGGGAAT 120

GCCTGTATCT GTTGCCCTTT TTGCTTCAGG TGTTGGAACA CTCATCTACA TGATTGCTAC 180

TGGTTTTAAA GTTCCAGTTT ATCTAGGTTC TTCATTTGCC TTTATCACAG CTATGTCACT 240

GGCTATGAAA GAAATGGGGG GGGATGTATC TGCTGCCCAA ACAGGGGTTA TCTTGACTGG 300

TTTGGTCTAT GTCCTTGTTG CTACCAGCAT CCGATTTGTA GGAACAAAAT GGATTGATAA 360

ACTCTTGCCA CCAATCATTA TCGGTCCTAT GATCATCGTT ATCGGTCTTG GACTTGCAGG 420

TTCAGCTGTT ACCAATGCAG GTCTTGTAGC AGACGGAAAT TGGAAAAATG CTCTGGTAGC 480

CGTTGTTACT TTCCTAATTG CTGCCTTTAT CAATACAAAA GGAAAAGGCT TCCTACGAAT 540

CATTCCATTC CTCTTTGCCA TTATCGGTGG TTACCTTTTC GCACTAACTC TTGGCTTGGT 600

TGACTTTACA CCAGTTCTTA AAGCCAACTG GTTCGAAATT CCTGGTTTCT ACTTGCCATT 660

TAGCACAGGT GGTGCCTTTA AAGAGTACAA TCTTTACTTT GGTCCAGAAG CCATCGCTAT 720

CTTGCCAATC GCTATCGTAA CAATTTCTGA ACATATCGGA GACCATACTG TTTTGGGTCA 780

AATCTGTGGT CGTCAATTCT TAAAAGAACC AGGTCTTCAC CGTACTCTTC TTGGTGACGG 840

TATCGCAACT TCTGTTTCTG CCTTCCTTGG TGGACCAGCC AATACAACTT ACGGAGAAAA 900

TACAGGGGTT ATCGGTATGA CTCGTATCGC TTCTGTCTCA GTTATCCGTA ACGCTGCCTT 960

CATCGCGATT GCCCTCAGCT TCCTTGGTAA ATTCACTGCC TTGATTTCAA CTATTCCAAA 1020

CGCTGTACTT GGTGGTATGT CAATCCTTCT CTATGGGGTT ATCGCCAGCA ATGGTTTGAA 1080

AGTCTTGATT AAAGAACGTG TTGATTTCGC TCAAATGCGA AACCTCATCA TCGCAAGTGC 1140

TATGTTGGTT CTTGGACTTG GAGGAGCTAT CCTTAAACTT GGTCCAGTTA CACTTTCAGG 1200

TACTGCCCTT TCAGCCATGA CAGGAATCAT CTTGAACTTG ATCTTGCCAT ACGAAAATAA 1260

AGACTAAGAG TCTAAATACA CCTAATCCAC TCAGACAGCT GAGTGGATTT TTCGTATACC 1320

ATAATAAAAG TGTCTTAACA AAATTATTAA AATCAAAAAA CGTATAATAT CAGATATTCT 1380

AAAACCTTGA TACTGTACGT TTTATCATAG AAATTTTTAC TTTATTTTCT CATCAAATGA 1440

GATTTGCATC AATCTCTTGT CTTACTTGCG TTTCTTCTTC GCTTTCTTCA TTTTGTTAGC 1500

CATACGTTTC ATGGACTGTT TCATGGCAAA TTCACCAATT TTACCTTTCA AACCGCCACC 1560

AAACATCTGG CTCATATCTG GCATTCCTGC TCCTCCGAGA GCTGATAAGT CAGGCATACC 1620

GCCTTGTCCC ATCATTCCTT CAAGGGCAGA CATATCCATT CCTCCCATAT TTGGCATATT 1680

TTTAGGAAGG TTATTTGGAT TAATCCCCAT TTGCTTCATC ATTTTATTCA TATCCCCAGA 1740

CATAACACCC TGCATGAGCT GTTTAGCCTG GTTAAAGTCC TTGATGAATT TATTGACTTC 1800

GACGAATGTA TTTCCAGAAC CAGCAGCAAT ACGACGGCGA CGGCTTGGAT TTAACAAATC 1860

TGGGTTTTCA CGCTCTTCAG GTGTCATCGA AGACACAATG GCACGTTTAC GAGCAATCTG 1920

GCGTTCATCC ACCTTCATGT TTTGAAGGGC TGGATTGTTG GCCATACCTG GAATCATCTT 1980

GAGCAAGTCT TCCATCGGCC CCATATTTTG CACCTGATCT AATTGATCGA TGAAATCATT 2040

AAAATCAAAG GTGTTTTCGC GCATCTTCTC AGCCATTTCA AGGGCTTTTT GTTCATCGTA 2100

TTCCTGAGAA GCTTTCTCAA TCAAAGTGAG CATATCCCCC ATACCAAGGA TACGGCTAGA 2160

CATGCGGTCT GGGTGGAAGG TTTCAATGTC CGTAATCTTT TCACCTGTAC CAGTGAACTT 2220

GATTGGTTTT CCAGTAATGT GACGAACAGA CAGAGCAGCA CCACCACGAG TATCGCCATC 2280

AATCTTGGTA AGGATGACCC CAGTCACTTC CAACTGAGCA TTAAACTCAC GCGCAACATT 2340

GGCTGCTTCC TGACCAATCA TAGCATCAAC GACAAGCAAG ATTTCATTTG GTTGAGCCAA 2400

TGCTTTCACA TCACGAAGCT CATTCATGAG GAGCTCATCA ATCTGCAAAC GACCCGCAGT 2460

ATCAATCAAG ACATAGTCGT TATGATTAGT TTGGGCTTGC TCCAAACCTT GACGTACAAT 2520

CTCAACAGCT GGTACTTCTG TTCCAAGTGC AAAGACAGGC ACATCAATCT GTTGTCCCAA 2580

GGTCTTAAGC TGGTCAATGG CAGCTGGACG ATAAATATCC GCCGCAATCA TCAAAGGACG 2640

AGCATTTTCT TCTTTCTTGA GTTTGTTGGC CAATTTACCA GCAAAGGTTG TTTTACCAGC 2700

CCCTTGTAAA CCAACCATCA TGATGATGGT TGGAATCTTA GGTGACTTGA TAATTTCTGC 2760

CGTATCAGAA CCTAAAACGG CTGTCAATTC CTCATCAACG ATTTTAATAA TCTGTTGCGC 2820

AGGATTAAGT GTATCAATGA CCTCATGCCC GACTGCACGC TCACGAACTT TCTTGATAAA 2880

GTCCTTTACA ACAGGCAAGG CAACGTCGGC CTCGAGCAAG GCCAAGCGAA TTTCTTTGGT 2940

TGCCTCTTGG ACATCAGATT CAGAGATTTT TCCTTTTTTA CGTAGATTTT TAAAGACGTT 3000

CTGCAAACGT TCTGTTAAAC TTTCAAATGC CATTTTTCTT CCTCTTATTC TCTATTATCA 3060

ATGCTTGTTA AAATTTCTAT CTGCTCCTGC AGAAAGTCAT CCTTGGGATA GCGCTCCAAA 3120

ATCTGATCAA AAATCTGACT GCGGACAATA TAGTCCGAGT ACATGTGCAA TTTCATCTCA 3180

TAATCTTCCA GAATCTTTTC TGTTCGCTTG ATATTGTCAT AGACAGCCTG ACGACTGACA 3240

CCGAACTCCT CGGCAATTTC AGCAAGGCTG TAATCATCAG CGTAGTAGAG CTCGATATAA 3300

TTCATTTGCT TATCTGTCAA AAGCGCCGCA TAAAATTCAA AGAGCGCATT CATACGATTG 3360

GTTTTTTCGA TTTCCATAAC TTTTATTATA CCAAAAATTA GCCTAATCTA CCACACTAGG 3420

AAGCCGATCC AAGAAGATAG ATAGCTAAAT TTGAAAAAGA CATGAGCCTA GCCCCAAGTA 3480

ATTTCCAATT GATAGCTGGC AAAGGGATGT CCCTCTTGAT TTTGTAGTTG ATAATCTAGT 3540

TCAATCTTTT GCCTATCAAC TTGATAATGG CTCGTTTGGA TGATAAACTC CTGCATGCCC 3600

ATAGGTGTAG GAATATAGGC TAAACTATCG CTATCCTTTA GAAAGCGCAT AATGGTCTTG 3660

GGATTAGAAA ATCGGCTCAT CACAAGTTCT TGACCATGAA ATTTAATCAC TACTTTTTCC 3720

TTTTCCTCAT TATAGAAAAG CAGGTAGCTA TAATCTCCTT TTTCATGCAC TTCCACATCA 3780

TAAAGCTGGT CAATCACTTC CAACTGCTCA TCAAACTGAA TCGTATTTCG CATCCGAATC 3840

TTCACATCAG GCCCTCTTTC TTGTCTCTTG TCCTACTATT TTACCAAAAA GAGCAGGATT 3900

TTGCTATAAT GGTCATATGA ACGAAAAAGT ATTCCGTGAC CCTGTTCACA ACTACATCCA 3960

TGTCAATAAT CAAATCATCT ATGACTTGAT TAATACAAAA GAATTTCAGC GTTTGCGCCG 4020

GATCAAACAA CTGGGAACTT CCAGTTATAC CTTCCACGGT GGAGAACACA GTCGCTTCTC 4080

TCACTGTCTA GGAGTCTATG AAATTGCACG ACGCATCACA GAGATTTTCG AAGAAAAATA 4140

TCCTGAGGAA TGGAATCCTG CCGAGTCTCT CTTGACCATG ACCGCTGCTC TCCTACACGA 4200

CCTTGGGCAT GGTGCCTACT CCCATACTTT TGAACATCTC TTTGATACAG ACCATGAAGC 4260

CATTACTCAG GAGATTATTC AAAATCCTGA GACAGAGATT CACCAAGTCC TGCTACAAGT 4320

GGCACCTGAT TTCCCAGAAA AGGTGGCCAG TGTCATTGAC CATACCTATC CTAATAAGCA 4380

GGTCGTGCAG CTCATTTCTA GTCAGATTGA CGCAGATCGC ATGGACTATC TCTTGCGCGA 4440

CTCCTATTTT ACAGGAGCAT CCTATGGGGA ATTTGACCTG ACTCGAATCC TCCGAGTCAT 4500

TCGTCCTATC GAAAATGGTA TCGCCTTTCA GCGCAATGGC ATGCACGCCA TCGAAGACTA 4560

CGTCCTCAGT CGCTACCAGA TGTACATGCA GGTTTATTTC CACCCCGCAA CACGCGCCAT 4620

GGAAGTTCTC CTACAGAATC TTCTCAAACG CGCCAAGGAA CTCTATCCTG AGGACAAGGA 4680

TTTCTTTGCC CGAACTTCTC CACACCTCCT GCCTTTCTTC GAAAAAAATG TGACCTTGAC 4740

TGACTATCTG GCTCTGGATG ATGGCGTGAT GAATACCTAC TTCCAGCTTT GGATGACCAG 4800

TCCTGACAAG ATTCTTGCAG ATTTATCGCA TCGCTTTGTC AACCGCAAGG TCTTTAAATC 4860

CATTACCTTT TCACAAGAGG ACCAAGATCA ACTTACTAGC ATGAGAAAAT TGGTTGAGGA 4920

TATCGGCTTT GATCCCGACT ACTACACTGC CATTCATAAG AACTTTGACC TCCCTTATGA 4980

TATCTATCGT CCCGAATCTG AAAACCCACG GACACAGATT GAGATTTTAC AAAAAAATGG 5040

AGAACTGGCC GAACTCTCTA GCCTGTCTCC TATCGTCCAA TCCCTTGCTG GCAGTCGCCA 5100

CGGAGATAAT CGCTTTTATT TTCCAAAAGA AATGTTGGAC CAAAACAGCA TCTTTGCAAG 5160

CATTACCCAG CAATTTTTAC ACTTGATTGA GAACGATCAT TTTACCCCAA ATAAAAACTA 5220

GAAGAGGAAA TTTATGAGTA TTAAACTAAT TGCCGTTGAT ATCGACGGAA CCCTTGTCAA 5280

CAGCCAAAAG GAAATCACTC CTGAAGTTTT TTCTGCCATC CAAGATGCCA AAGAAGCTGG 5340

TGTCAAAGTC GTGATTGCAA CTGGCCGCCC TATCGCAGGC GTTGCCAAAC TTCTAGACGA 5400

CTTGCAGTTG AGAGACGAGG GGGACTATGT GGTAACCTTC AACGGTGCCC TTGTCCAAGA 5460

AACTGCTACA GGACATGAGA TTATCAGCGA ATCCTTGACT TATGAGGATT ATCTAGATAT 5520

GGAATTCCTC AGTCGCAAGC TCGGTGTCCA CATGCATGCC ATTACCAAGG ACGGTATCTA 5580

TACTGCAAAT CGCAATATCG GAAAATACAC TGTACACGAA TCAACCCTCG TCAGCATGCC 5640

TATCTTCTAC CGTACCCCTG AAGAAATGGC TGGCAAAGAA ATTGTTAAAT GTATGTTTAT 5700

CGATGAACCA GAAATTCTCG ATGCTGCGAT TGAAAAAATT CCAGCAGAAT TTTACGAGCG 5760

CTACTCCATC AACAAATCTG CTCCTTTCTA CCTCGAACTC CTTAAAAAGA ATGTAGACAA 5820

GGGTTCAGCC ATTACTCACT TGGCTGAAAA ACTCGGATTG ACCAAAGATG AAACCATGGC 5880

AATCGGTGAT GAAGAAAATG ACCGTGCCAT GCTGGAAGTC GTTGGAAACC CCGTTGTCAT 5940

GGAAAATGGA AATCCAGAAA TCAAAAAAAT CGCCAAATAC ATCACCAAAA CAAATGACGA 6000

ATCCGGCGTT GCCCATGCCA TCCGAACATG GGTACTGTAA AAGTATCATT TTTCAATAAG 6060

AATTGATTAG CAATAAAATC CAATGAATTT TTTTAGCAAA CTATTTAATT TAAAACAAAA 6120

TAATCATAAT AGAGACACAA ATTCTGATTG TAACAATTTT TACCTAAACG AATTAGAATG 6180

TGGCCTTACT CCTGGGCAAC TCATACTCAT AGATTGGACT CAAAAAACAG GGAGAAATTA 6240

TAATTTCCCA AGATATTTTA AATACTCTCT TCAAATTGAC CCTGAATCTA CACACAATCA 6300

ATTATACAAA TTAGGATACT TCACTAAAAA TAAGACTTTA TCATATCTTA CAGTAGTAGA 6360

ATTAAAAACT ATATTATCTA AACATAATTT AGCTACTTCT GGAAAAAAAG CAGAATTAAT 6420

TACAAGAATA ATTAATAATG TTAACATTGA CAATTTAGAT ATTCCGTTCG AATTTAAACT 6480

AACAAAAGAA GCACAAAATC TTATTATCGA ACATAGTGAC TATATCAAAG CATACTATGA 6540

TAAAGACATA ACTATGGAAG ATTATTGTAA AGAAAAAAAC AATATCTCTT TTAAAGCAAC 6600

TTTTGGTGAT ATAAAATGGA GTCTCTTAAA TAAACAAGCT CATAGGAATA CTGTATCAGG 6660

AGATTTTGGA TGCTTATCTA ACACACGAAA GGCTCAGGGA AGACATTTGG AACAAGAAGG 6720

TAATATTAAA CATGCTTTAA TATATTACAT AGAATCTTTG ATAATTACTA TTTCAGGATT 6780

AGAAAACAAT TTTTCAGCCA CTGATTATCC AGTATATTAT CCCGATTCGA TACCTGACTA 6840

CTCACTAAAA CATATTCAAA CATTAATGGA ATCATTATCT GATGACGATT ATGATTTTGC 6900

TTTTGATGAA GCATTATTTC GCTTCTCAAT TTTGAATGCA AATCATTTTT TATCTAAGGA 6960

AGATATTGAC TATTTAAGAG TTAATTTACC TCGTTCCACT GCTGAAGAAA TAAACAATTA 7020

CTTAAAGAAA TATGAATGTT ATAGTCCTTT AAATAATTTA GAACTTGACG ATTTTGAATA 7080

AATTGACTAT ACAAACATTT ATATACTCGA TATAGTCTCA ATTTTATCTG ATGATTGCCC 7140

AAATTTTTCA ATAATAAAAC GCATAATATT ATGGAGACAA TCCCCTATAT TATGCGTTCT 7200

TTTAATATCA AAGACTTTTT GACAAACTTC TTTGATATCT AATTACATGC CCCCTGCAGG 7260

AATCGAACCT GCAACTACTC CTTAGGAGGG AGTTGTTATA TCCATTGAAC TAAGGGAGCT 7320

AGATAAAAAC TCTGCTAAAT GAGCAGAGTT TTTTAGTCGA ATTAACGACG GATTTCTTTG 7380

ATACGAGCTG CTTTACCTTG AAGAGCACGC AAGTAGTACA ATTTCGCACG ACGTACTTTA 7440

CCGTAACGAA CAACTTCGAT TTTTTCAACA CGTGGAGTGT GGATTGGGAA GATACGCTCA 7500

ACACCTACAC CGTTAGAGAT TTTACGAACT GTGTAGTTTT CTGAGATTCC AGCACCTTTA 7560

CGTGCGATAA CAACACG 7577

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4945 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

CCTCGCTGAT GATTGGTGCT GTTTTATTTG CTGGTCCAGC CTTGGCTGAA GAAACTGCAG 60

TTCCTGAAAA TAGCGGAnCT AATACAGAGC TTGTTTCAGG AGAGAGTGAG CATTCGACCA 120

ATGAAGCTGA TAAGCAGAAT GAAGGGGAAC ATGCTAGAGA AAACAAGCTA GAAAAGGCAG 180

AAGGAGTAGC GATAGCATCT GAAACTGCTT CGCCAGCAAG CAATGAAGCT GCAACTACTG 240

AAACTGCAGA AGCAGCTAGC GCAGCTAAAC CAGAGGAAAA AGCAAGTGAG GTGGTTGCAG 300

AAACACCATC TGCAGAAGCA AAACCTAAGT CTGACAAGGA AACAGAAGCA AAGCCCGAAG 360

CAACTAACCA AGGGGATGAG TCTAAACCAG CAGCAGAAGC TAATAAGACT GAAAAAGAAG 420

TCCAGCCAGA TGTCCCTAAA AATACAGAAA AAACATTAAA ACCAAAGGAA ATCAAATTTA 480

ATTCTTGGGA AGAATTGTTA AAATGGGAAC CAGGTGCTCG TGAAGATGAT GCTATTAACC 540

GCGGATCTGT TGTCCTCGCT TCACGTCGGA CAGGTCATTT AGTCAATGAA AAAGCTAGCA 600

AGGAAGCAAA AGTTCAAGCC TTATCAAACA CCAATTCTAA AGCAAAAGAC CATGCTTCTG 660

TTGGTGGAGA AGAGTTCAAG GCCTATGCTT TTGACTATTG GCAATATCTA GATTCAATGG 720

TCTTCTGGGA AGGTCTCGTA CCAACTCCTG ACGTTATTGA TGCAGGTCAC CGTAACGGGG 780

TTCCTGTATA CGGTACACTC TTCTTCAACT GGTCTAATAG TATTGCAGAT CAAGAAAGAT 840

TTGCTGAAGC TTTGAAGCAA GACGCAGATG GTAGCTTCCC AATTGCCCGT AAATTGGTAG 900

ACATGGCCAA GTATTATGGC TATGATGGCT ATTTCATCAA CCAAGAAACA ACTGGAGATT 960

TGGTTAAACC TCTTGGAGAA AAGATGCGCC AGTTTATGCT CTATAGCAAG GAATATGCTG 1020

CTAAGGTAAA CCATCCAATC AAGTATTCTT GGTACGATGC CATGACCTAT AACTATGGAC 1080

GTTATCATCA AGATGGTTTG GGAGAATACA ACTACCAATT CATGCAACCA GAAGGAGATA 1140

AGGTTCCGGC AGATAACTTC TTTGCTAACT TTAACTGGGA TAAGGCTAAA AATGATTACA 1200

CTATTGCAAC TGCCAACTGG ATTGGTCGTA ATCCTTATGA TGTATTTGCA GGTTTGGAAT 1260

TGCAACAGGG TGGTTCCTAC AAGACAAAGG TTAAGTGGAA TGACATTTTA GACGAAAATC 1320

GGAAATTGCG CCTTTCTCTT GGTTTATTTG CCCCAGATAC CATTACAAGT TTAGGAAAAA 1380

CTGGTGAAGA TTATCATAAA AATGAAGATA TCTTCTTTAC AGGTTATCAA GGAGACCCTA 1440

CTGGCCAAAA ACCAGGTGAC AAAGATTGGT ATGGTATTGC TAACCTAGTT GCGGACCGTA 1500

CGCCAGCGGT AGGTAATACT TTTACTACTT CTTTTAATAC AGGTCATGGT AAAAAATGGT 1560

TCGTAGATGG TAAGGTTTCT AAGGATTCTG AGTGGAATTA TCGTTCAGTA TCAGGTGTTC 1620

TTCCAACATG GCGCTGGTGG CAGACTTCAA CAGGGGAAAA ACTTCGTGCA GAATATGATT 1680

TTACAGATGC CTATAATGGC GGAAATTCCC TTAAATTCTC TGGTGATGTA GCCGGTAAGA 1740

CAGATCAGGA TGTGAGACTT TATTCTACTA AGTTAGAAGT AACTGAGAAG ACCAAACTTC 1800

GTGTTGCCCA CAAGGGAGGA AAAGGTTCTA AAGTTTATAT GGCATTCTCT ACAACTCCAG 1860

ACTACAAATT CGATGATGCA GATGCATGGA AAGAGCTAAC CCTTTCTGAC AACTGGACAA 1920

ATGAAGAATT TGATCTTAGC TCACTAGCGG GTAAAACCAT CTATGCAGTC AAACTATTTT 1980

TCGAGCATGA AGGTGCTGTA AAAGATTATC AGTTTAACCT AGGACAATTA ACTATCTCGG 2040

ACAATCACCA AGAGCCACAA TCGCCGACAA GCTTTTCTGT AGTGAAACAA TCTCTTAAAA 2100

ATGCCCAAGA AGCGGAAGCA GTTGTGCAAT TTAAAGGCAA CAAGGATGCA GATTTCTATG 2160

AAGTTTATGA AAAAGATGGA GACAGCTGGA AATTACTAAC TGGCTCATCT TCTACAACTA 2220

TTTATCTACC AAAAGTTAGC CGCTCAGCAA GTGCTCAGGG TACAACTCAA GAACTGAAGG 2280

TTGTAGCAGT CGGTAAAAAT GGAGTTCGTT CAGAAGCTGC AACCACAACC TTTGATTGGG 2340

GTATGACTGT AAAAGATACC AGCCTACCAA AACCACTAGC TGAAAATATC GTTCCAGGTG 2400

CAACAGTTAT TGATAGTACT TTCCCTAAGA CTGAAGGTGG AGAAGGTATT GAAGGTATGT 2460

TGAACGGTAC CATTACTAGC TTGTCAGATA AATGGTCTTC AGCTCAGTTG AGTGGTAGTG 2520

TGGATATTCG TTTGACCAAG CCACGTACCG TTGTTAGATG GGTCATGGAT CATGCACGAG 2580

CTGGTGGTGA GTCTGTTAAC GATGGCTTGA TGAACACTAA AGACTTTGAC CTTTATTATA 2640

AAGATGCAGA TGGTGAGTGG AAGCTAGCTA AGGAAGTCCG TGGTAACAAA GCACACGTGA 2700

CAGATATCAC TCTTGATAAA CCAATCACTG CTCAAGACTG GCGCTTGAAT GTTGTCACTT 2760

CTGACAATGG AACTCCATGG AAGGCTATTC GTATCTATAA CTGGAAAATG TATGAAAAGC 2820

TTGATACTGA GAGTGTCAAT ATTCCGATGG CCAAGGCTGC AGCCCGTTCT CTAGGCAATA 2880

ACAAGGTACA AGTTGGCTTT GCAGATGTAC CGGCTGGAGC AACTATTACC GTTTATGATA 2940

ATCCAAATTC TCAAACTCCG CTCGCAACCT TGAAGAGCGA AGTTGGAGGA GACCTAGCAA 3000

GTGCACCATT GGATTTGACA AATCAATCTG GTCTTCTTTA TTATCGTACC CAGTTGCCAG 3060

GCAAGGAAAT TAGTAATGTC CTAGCAGTTT CCGTTCCAAA AGATGACAGA AGAATCAAGT 3120

CAGTCAGCCT AGAAACAGGA CCTAAGAAAA CAAGCTACGC CGAAGGGGAG GATTTGGACC 3180

TTAGAGGTGG TGTTCTTCGA GTTCAGTATG AAGGAGGAAC TGAGGACGAA CTCATTCGCC 3240

TAACTCACGC AGGTGTATCA GTATCAGGTT TTGATACGCA TCATAAGGGA GAACAGAATC 3300

TTACTCTCCA ATATTTGGGA CAACCGGTAA ATGCTAATTT GTCAGTGACT GTCACTGGCC 3360

AAGACGAAGC AAGTCCGAAA ACTATTTTGG GAATTGAAGT AAGTCTGGAA CCGAAAAAAG 3420

ATTACCTAGT TGGTGATAGC TTAGACTTGT CTGAAGGACG CTTTGCAGTG GCTTATAGCA 3480

ATGACACCAT GGAAGAACAT TCCTTTACTG ATGAGGGAGT TGAAATTTCT GGTTACGATG 3540

CTCAAAAGAC TGGTCGTCAA ACCTTGACGC TTCATTACCA AGGCCATGAA GTTAGCTTTG 3600

ATGTTTTGGT ATCTCCAAAA GCAGCATTGA ACGATGAGTA CCTCAAACAA AAATTAGCAG 3660

AAGTTGAAGC TGCTAAGAAC AAGGTGGTCT ATAACTTTGC TTCATCAGAA GTAAAAGAAG 3720

CCTTCTTGAA AGCAATTGAA GCGGCCGAAC AAGTGTTGAA AGACCATGAA ACTAGCACCC 3780

AAGATCAAGT CAATGACCGA CTTAATAAAT TGACAGAAGC TCATAAAGCT CTGAATGGTC 3840

AAGAGAAATT TACGGAAGAA AAGACAGAGC TTGATCGCTT AACAGGTGAG GTTCAAGAAC 3900

TCTTGGCTGC CAAACCAAAC CATCCTTCAG GTTCTGCCCT AGCTCCGCTT CTTGAGAAAA 3960

ACAAGGCCTT GGTTGAAAAA GTAGATTTGA GTCCAGAAGA GCTTACAACA GCGAAACAGA 4020

GTCTAAAAGA TCTGGTTGCT TTATTGAAAG AAGACAAGCC AGCAGTCTTT TCTGATAGTA 4080

AAACAGGTGT TGAAGTACAC TTCTCAAATA AAGAGAAGAC TGTCATCAAG GGTTTGAAAG 4140

TAGAGCGTGT TCAAGCAAGT GCTGAAGAGA AGAAATACTT TGCTGGAGAA GATGCTCATG 4200

TCTTTGAAAT AGAAGGTTTG GATGAAAAAG GTCAAGATGT TGATCTCTCT TATGCTTCTA 4260

TTGTGAAAAT CCCAATTGAA AAAGATAAGA AAGTTAAGAA AGTATTTTTC TTACCTGAAG 4320

GCAAAGAGGC AGTAGAATTG GCTTTTGAAC AAACGGATAG TCATGTTATC TTTACAGCAC 4380

CTCACTTTAC TCATTATGCC TTTGTTTATG AATCTGCTGA AAAACCACAA CCTGCTAAAC 4440

CAGCACCACA AAACACAGTC CTTCCAAAAC CTACTTATCA ACCGACTTCT GATCAACAAA 4500

AGGCTCCTAA ATTGGAAGTT CAAGAGGAAA AGGTTGCCTT TCATCGTCAA GAGCATGAAA 4560

ATACTGAGAT GCTAGTTGGG GAACAACGAG TCATCATACA GGGACGAGAT GGACTGTTAA 4620

GACATGTCTT TGAAGTTGAT GAAAACGGTC AGCGTCGTCT TCGTTCAACA GAAGTCATCC 4680

AAGAAGCGAT TCCAGAAATT GTTGAAATTG GAACAAAAGT AAAAACAGTA CCAGCAGTAG 4740

TAGCTACACA GGAAAAACCA GCTCAAAATA CAGCAGTTAA ATCAGAAGAA GCAAGCAAAC 4800

AATTGCCAAA TACAGGAACA GCTGATGCTA ATGAAGCCCT AATAGCAGGC TTAGCCAGCC 4860

TTGGTCTTGC TAGTTTAGCC TTGACCTTGA GACGGAAAAG AGAAGATAAA GATTAAATAT 4920

CGAAAAATCT TGTGAAATCT TTCCG 4945

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25002 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

GACAACTCAA GTAGCTTTTT CTTATTTTGA AAAAGGAGAT CAGAGTTTAA CTATGTCAGA 60

AAAATCACAA TGGGGGTCGA AACTTGGTTT TATTCTAGCA TCTGCTGGCT GGCCATCGGG 120

CTTGGTTCCG TTTGGAAGTT TCCCTACATG ACTGCTGCTA ATGGCGGTGG AGGCTTTTTA 180

CTAATCTTTC TCATTTCCAC TATTTTAATC GGTTTCCCTC TCCTGCTGGC TGAGTTTGCC 240

CTTGGCCGTA GTGCTGGCGT TTCCGCTATC AAAACCTTTG GAAAACTGGG CAAGAATAAC 300

AAGTACAACT TTATCGGTTG GATTGGCGCC TTTGCCCTCT TTATCCTCTT ATCTTTTTAC 360

AGTGTTATCG GAGGATGGAT TCTAGTCTAT CTAGGTATTG AGTTTGGGAA ATTGTTCCAA 420

CTTGGTGGAA CGGGTGATTA TGCTCAGTTA TTTACTTCAA TCATTTCAAA TCCAGCCATT 480

GCCCTAGGAG CTCAAGCGGC CTTTATCCTA TTGAATATCT TCATTGTATC ACGTGGGGTT 540

CAAAAAGGGA TTGAAAGAGC TTCGAAAGTC ATGATGCCCC TGCTCTTTAT CGTCTTTGTT 600

TTTATCATCG GTCGCTCTCT CAGTTTGCCA AATGCCATGG AAGGGGTTCT TTACTTCCTC 660

AAACCAGACT TTTCAAAACT GACTAGCACT GGTCTCCTCT ATGCTCTGGG ACAATCTTTC 720

TTTGCCCTCT CACTAGGGGT TACAGTCATG TTGACCTATG CTTCTTACTT AGACAAGAAA 780

ACCAATCTAG TCCAGTCAGG AATCTCCATC GTAGCCATGA ATATCTCGAT ATCCATCATG 840

GCAGGTCTAG CCATTTTCCA AGCTCGATCC CCCTTCAATA TCCAGTCTGA AGGGGGACCC 900

AGCCTGCTCT TTATCGTCTT GCCTCAACTC TTTGACAAGA TGCCTTTTGG AACCATTTTC 960

TACGTCCTCT TCCTCTTGCT CTTCCTTTTT GCGACAGTCA CTTTTTCTGT CGTGATGCTG 1020

GAAATCAATG TAGACAATAT CACCAACCAG GATAACAGCA AACGTGCCAA ATGGAGTGTT 1080

ATTTTAGGAA TTTTGACCTT TGTCTTTGGC ATTCCTTCAG CCCTATCTTA CGGTGTCATG 1140

GCGGATGTTC ACATTTTTGG TAAGACCTTC TTTGACGCTA TGGACTTCTT GGTTTCCAAT 1200

CTCCTCATGC CATTTGGAGC TCTCTACCTT TCACTTTTTA CAGGCTATAT CTTTAAAAAG 1260

GCTCTTGCAA TGGAGGAACT CCATCTCGAT GAAAGAGCAT GGAAACAAGG ACTGTTCCAA 1320

GTCTGGCTCT TCCTTCTTCG TTTCTTCGTT TCGTCATTCC AATCATCATC ATTGTGGTCT 1380

TCATTGCCCA ATTTATGTAA TCAAAAAGGA CTTGAGTAGT GAACTCAGGC CCTTTCTTTT 1440

TATGGATGGC TAACAATCAA TTCCAAACCT TGCCCTTCCA GAGTCCAAGC TTCAACATCA 1500

CTTGGTAGGA TAAAGTGGCT GCCTTTTTGA ATTGGATAAT TTTTCCCGTC AACAGTTAGC 1560

TGACCTTGAC CAGCCAAGAC ACTCAATAAG CTGTAGTCAG CTGTCTTTTC AAAGTCAACT 1620

TTTCCAGTAA TTTCCCACTT GTAAACTGCG AAGAAATCAT TAGATACAAG GAGAGTGGAA 1680

CGCAAATCAT CTGCTTTAAC AGTTACAGGA CGGCTATTTG CTGGCTCACC AATGTTCAAG 1740

ACATCGATGG ATTTTTCAAG ATGAAGTTCA CGCAAGTTGC CTTTGTCATC CTTGCGGTCA 1800

AAGTCATAGA CGCGATAGGT GGTATCGCTA GACTGCTGGG TTTCAAGGAT TAAGATACCC 1860

GCCCCGATAG CGTGCATAGT CCCGCTTGGT ACATAGAAGA AATCTCCAGC CTTAACAGGG 1920

ACTTTGGTCA ACAAGTCATC CCAGTTCTTG TCCTCGATTT GCTGGCGGAG TTCTTCTTTT 1980

GACTTGGCAT TGTGACCGTA GATAATCTCT GAACCTTCAT CCGCTGCGAT AATGTACCAG 2040

CATTCTGTTT TTCCGAGTTC GCCTTCATGC TCGAGTCCAT AAGCATCGTC TGGGTGAACT 2100

TGGACACTGA GCCAGTCGTT GGCATCGAGG ATCTTGGTCA AAAGTGGAAA TACAGGTTCT 2160

GGACGATTGC CAAATAATTC ACGGTGTTCC GCATACAAAG TAGCAAGATC TGTTCCCTCG 2220

TAACGACCAT TGGCAACTTT AGAGACTCCA TTTGGATGGG CTGAGATGGC CCAATATTCT 2280

CCGATTTTTT CACTTGGGAT GTCGTAGCCA AACTCATCAC GTAGCTTGGC TCCACCCCAG 2340

ATTTTTTCTT GCATAACTGA TTGTAAAAAT AATGGTTCTG ACATGTCGAT CTCCTGTCTG 2400

ATTTTTCTCC CCTCATTATA GCAAAAAAAG AGTTCGAATT GAACTCTTTT TTACATCTTA 2460

TAAAGCAGGG AGAAGATTTT ATAAAAATAG TAAACAAATG TGCTCTACCC GATGCTTGCA 2520

CCATTGCTAT AAATGACATC CTTGTACCAA TAGAAGGACT TCTTCTTGCT ACGTTTGAGA 2580

GCTCCGTTTC CTACATTATC TCGATCTACA TAGATAAAGC CATAGCGCTT ATTCATTTCC 2640

CCTGTGCCAG CTGAAACCGG ATCGATACAG CCCCAAGTCG TATAACCAAG CAAGTCAACC 2700

CCGTCTTGGT AAATGGCATC TCGCATGGCC TTGATGTGGG CCTCTAAGTA AGTAATCCGA 2760

TAGTCATCTG CTACATAACC ATTCTCATCC GGTGTATCCA TAGCACCGAG TCCATTTTCT 2820

ACGATAATAC TAAACTAAAA TCAAAAAGCA TTATATAATA GTGATATGAA ATCAACTAAA 2880

GAAGAAATCC AAACCATCAA AACACTTTTA AAAGACTCTC GTACAGCTAA ATATCATAAA 2940

CGCCTTCAAA TCGTTCTATA GTAAAATGAA ATAAGAACAG TACAAATCGA TCAGGACAGT 3000

CAAATCGATT TCTAACAATG TTTTAGAAGT AGGGGTGTAC TATTCTAGTT TCAATCTACT 3060

ATATTTCGTC TGATGGGCAA ATCTTATAAA GAGATTATAG AACTTTTATA GTAGTTTGAA 3120

ATAAGATGTG AACAACTCTA TCAGGAAAGT CAAATTAATT TATAGAAATA TTTTAGCAGC 3180

CAAGGTGTAC TGTTATAGAT TCAATACACT ATAGACTGTA ATCAAACAAC GATTTGGCGA 3240

AATGTAAAAA AATATGAGGA GTTCGGACTC GACTCTCTCC TTCAAGAAAC ACGTGGTGGT 3300

CGTAACCATG CATATATGAC AGTTGAGGAA GAGAAAGCCT TTCTTGCCCG CCATTTGAAG 3360

GCTACAGAGG CAGGAGAATT TGTTACAATT GATGCCTTAT TTCAGGCTTA TAAAAAGGAG 3420

TTAGGTCGTT CCTACACACG TGATGCCTTC TATCAACTGT TGAAGCGCCA TGGTTGGCGA 3480

AATATTACGC CACGTCCAGA ACATCCTAAG AAAGCAGACG CTCAAACCAT TGTTGCGTCT 3540

AAAAATAAAA TCTCAATCCA AGAAGGCAAG AAAGCGTTTT AAATATAGTA GACGTTTTCG 3600

TAAGGTTTGC TTGATGTACC AAGCTGAAGC TGGTTTCGGT AGAATCAGTA AACTGGGATC 3660

TTGTTGGGCT CCAATAGGAG TAGGTCCACA TATCCATAGT CACTATATAC GAGAATTTCG 3720

CTATTGTTAT GGAGCTGTTG ATGCCTATAC AGGCGAATCA TTTTTCTTAA TAGCTGGTAG 3780

ATGTAATACT GAGTGGATGA ACGCCTTTTT AGAAGAGCTT TCACAAGCTT ATCCTTTTAC 3840

TCGTTATGGA CAATGCTATA TGGCATAAAT CAAGTACCTT AAAGATTCCG ACTAATATTG 3900

GTTTTGCATT TATTCCTCCA TACACACCAG AGATGAACCC CATTGAACAA GTGTGGAAAG 3960

AGATTCGTAA ACGTGGATTT AAGAATAAAG CCTTTCGAAT TTTGGAAGAT GTCATGAATC 4020

AACTCCAAGA TGTCATACAA GGATTGGAGA AGGAGGTGAT AAAGTCCATC GTTAATCGGA 4080

GATGGACTAG AATGCTTTTT GAAAGCAGAT GAGTATTATA TGCAATTTCT TTATATAAAA 4140

AGACCGGATT GCTCCGATCT TTCAATAGTT CATATTCTCA ATTTCTATTT TAAAAATAGC 4200

TAAGGTTAAC GTCAAATGAC TACGCGACCT ATTTCATACG ATAAAAATCA AGCACTAGAC 4260

CAGCAGGTCC TTGAACTAAT AAGGACTCTG TTCCCCAATC GGTTACAGTT GGTCCGTGTA 4320

AAACCTTTAT ACCAAGCTCG TTCAACCGTT TGTAGTTCTG GTCTACATCC TCAACCTCGA 4380

TATGAATAAT GATTCCTGAC TGAAAGTTTT CCAAAGGAAC CAAATGATTT TGTGACAACA 4440

TAAGGCAGTG ACTACCAATC GTAAACTGAG CAAAACCATC ATTAGCATAA TCTGCCTTTT 4500

TATCCAAGAT ATGCTCCAAG TCAGCACAGA CTTGGGGAAC ATTTGAAACG ATAATATCTA 4560

ATTGATTTAA ATTCATTTAC TCTCCTCCAT AAAAAGACCG GATTGCTCCG ATCTTTTAAA 4620

GTTCTGCTCT ATGAAAATCA AAGAATAAAG TCTACAAGTT TCATATTTGA TTTTCGGCGA 4680

GAGGAATTAT TTAATTGCGC GTGATTGCAA TCCTTCTTCT TCCAAGAAGA GACGGAATGG 4740

TACGAGTTCT TCTGCTTCGT ATTTTTCCTT GAAGGCTTTG ATAGCTTCTT CTGAGTGAAG 4800

TTTTGGATCC AATTCAAGTA CTTCTACTGG AAGTGGACGG TGTTGAGTGA TGCGAGCATC 4860

GATGACAACA GTTTTACCTT CTTTGTTCAA TTTAACAGCT TCTGCAACAA CTGCATCGAT 4920

GTCTTCGATA CGGTCAACTG TGAATCCAAC AGCTCCTTGA GCTTCCGCAA TTTTAGCGTA 4980

GTCAGCGTTT GTGAAGTCTA CACCAAACAA GTGTTTGTTT GTATCTTCGT ATTTGTTCTT 5040

GATGAAGCCG TACTCAGCAT TTGAGAAGAC AAGGTTGATA ACTGGAAGGT CGTATTGAAC 5100

GTTTGTGATA ACGTCTGGGT AGCACATGTT GAATGCTCCG TCACCCATGA TGTTCCATAC 5160

TTGGCGATCT GGATTGTCTT TCTTAGCAGC GATACCACCA GGAAGGGCAA TACCCATTGT 5220

CGCAAAGAGT GGAGATGTAC GCCACATGTT CTTAGGTGTC ATGTGAAGGT GACGAGTAGA 5280

TGTTTGAGTA GTGTTACCTA CGTCGATTGA GTAGATAGCG TCTTGATCAG CATGTTTGTT 5340

GATTGCATTG TAAACTTGAT ACAATTGCAA TTCACCCTCA GTTTTACCTT CGAGTTTGTT 5400

CATGTAATCA CGCCAGTTTT GGTTGTTCTT AACGTTTGCA CGCCACCATG GAGTTGATTC 5460

AACTGGGTTT ACTTTGTCAA GGATAGCTTT AGCTGCTTGA CCAGCATCAC CAAGGATTGA 5520

AGCGTCAAGG GCATGACGTT TACCAAGTTT GTAAGGGTCG ATATCGACTT GGATGAATTT 5580

TTCAGTGTTC TTGAATGCTT CGTAAACTTC AGCAAATGGG AAGTTTGAAC CAAGGAAAAG 5640

AACTGTGTCT GCTTCAAAGA CCACTTCGTT GGCTGGTTTC CAACCAACAC GGTAAGCAGA 5700

ACCTGTCAAA CCTTCATAGT TCCATTCGAA AGCTTCAAAG TTTTTACCAG TTGTGATGAT 5760

TGGTGCTTTG ATTTTACGTG ACAATTCAGT AATCACTTCA CCAGCTTTAA CACCACCAAA 5820

TCCAGCATAG ATAACTGGGC GTTCAGCATT GTTCAAGATT TCAACAGCTT TGTCGATTTC 5880

AACTTCGTTC AAAGCAGGAG CGATGAATGA GCGTTCGTAT GAACCTGAAC CGTAGTATGA 5940

GTTTTCATCG ATTTCTTGGA AACCGAAGTT TACTGGAATT TCAACAACAG CTGGACCTTT 6000

TTTAGAAACT GCAGCACGGC AGGCTTCGTC AATTACTTTT GGCAATTGCT CAGCGTAAGC 6060

TACACGTTTG TTGTAAACAG CGATACCGTT GTACATTGGG TTTTGGTTAA GCTCTTGGAA 6120

AGCATCCATG TTCAATTCGT TAACTGGACG TGATCCAAGG ATCGCTAGGA ATGGAGTGTT 6180

ATCCATAGCT GCATCGTAAA CACCGTTAAT CAAGTGAGTC GCACCTGGAC CACCTGAACC 6240

AACTGCAACC CCGATTGAGC CGCCGAATTT AGCTTGCATA ACCGCTGCAA GAGCACCTGT 6300

CTCTTCGTGG CGAACTTGTA AGAAACGGAT ATCTTTGTCT TCAGCCAAAG CGTCCATCAA 6360

TGAGCTGAGT GTTCCTGATG GGATACCGTA GATTGTATCT ACGCCCCATG TTTTCAATAC 6420

GTTAAGCATT GCTGCAGATG CAGTAATTTT CCCTTGAGTC ATAATGATAA CTCTCCTTCA 6480

ATTTTTTTAA ACTTGGAGAA TACGATTACA TAGAATTGGA AACGTTCTCC AAATTTTTAC 6540

TATTCCACTG TATCATATTT ATGCTGACTT TTCTAAAAAT CTGCTCAAAA CTCTCTATTC 6600

TCTATTCTAA TACAGTTTTG AAAGTTCTGT CATTTCTGTT TTATAACAAA GAAATCTAGT 6660

CATTACTTTT AGTCTATTTT ACTAAAATTT AACAGAAGGG AACTGGTCAG AACAGATACA 6720

GAACTAAAGG CCATGGCTAG ACCTGCCAAT TCTGGGTTGA GAGCCAGTCC AACACCTGAA 6780

AAGACTCCTG CTGCAATCGG AATTCCGACA ACATTGTAGA TAAAAGCCCA GAAAAGATTG 6840

AGTAGAATTC GATGAAAGGT TTTCTTACTC ATATCAAAGG CACGAACCAC TCCTAAAAGA 6900

TTATTGGTTG TCAACACCAA ATCTGCTGAC TCGATGGCGA TATCTGTTCC AGCTCCCATA 6960

GCAATCCCCA CATCTGCTAC ACTAAGGGCA GGAGCGTCAT TGATACCGTC CCCAACAAAG 7020

GCTACTTTCC CTGACTGTTG CAGTTTATGG ATTTCATGGG CTTTTTCTTC TGGCAAGACG 7080

CCTGCAATGA CCTCTTCAAT TCCGATTTGA TCTGCAATAG CACGCGCCAC ACCAGCATTG 7140

TCTCCTGTCA GCATGACTGT TCGGAGACCA CGTTTTTTTA GCTGACTGAT GGCTAGCTTA 7200

GCATTTTCCT TAGGAATATC TTGCAAAGCA AGCAAGCCTT TGATTTCATT GTCAACAGCT 7260

AAGAACACAA CTGTCTTAGC TTCTTTTTCT AGTTCTTCTA GTTTATCTTG ATAAGTATTA 7320

GAAATATCCA TGCCATCCAG CATTTTAGCA TTTCCAAGTA AAACTTGTTT TCCATTGATT 7380

CGCCCTGAAA CACCTTTCCC GTGCAAGGAC TGAAAATTTT CAACAGTTTG AAACTCAAGT 7440

CCAGCTTCAC TCGCTCGCTT AACGATAGCC TCAGCCAGTG GGTGTTGAGA AGCATCTTCC 7500

AAGGAGGCTG CCAACCCAAA CACTTCTACT TCGTCGCCGA TGACATCTGT TACCACAGGT 7560

TTCCCTTCCG TCAAAGTCCC GGTCTTATCA AAGACAAGGG TTTGAACTTT CTGGATTTCC 7620

TGTAAGACAG TTCCATTTTT GAGGAGAACC CCCATCTTGG CACTACGTCC TGTCCCCACC 7680

ATAAGGGCTG TCGGTGTTGC AAGTCCCAAG GCACAAGGAC AGGCGATAAT CAAAACCGCC 7740

ACTCCGTAGA GAAGAGAGGA CACAAAGCTA GCTCCAAGCA CAACCACACT ATCCCTGAGC 7800

AAGACGAACC AAACCCAAAA GGTCATGATT CCTAAAATGA CAACTACTGG GACAAAAATC 7860

CCTGAAATCT TATCCGTCAA GTCCTGAATC GGCGCACGAC TTGTCTGAGC TTTCTTCACA 7920

AAATCCACAA TCTGAGCCAA AACAGTCTCT GAGCCAACTT TTTCTGCTCT AAAGACAAGC 7980

GTTCCACTAT GATTGATGGT TGAGCCAATG ACAGTATCTC CAACTGTCTT GTCCACAGGC 8040

AGACTCTCAC CTGTCACCAT GGATTCGTCA ATACTAGAGA CACCTTCTAC TACGACACCA 8100

TCAACAGCAA TCTTTTCACC GGGACGCACT CGAATCAGGT CGCCTACCTT GACTTGTTCC 8160

AAAGGAACTT GGACATAACT ATCATCACTC AAGACTTCTG CGGTTTTAGC TTGCAAGTCC 8220

AGTAATTTCT CCACAGCTTG GGACGTATTT TTTCTCATTT TTTCCTCAAA AACTGCTCCC 8280

AAAAGAACGA AAAAGAGGAT AAATCCAGCA CTTTCGAAGT AAACAGGGAG ACCAGCAAAG 8340

AGAGCAACTA GGCTATAGAA ATAAGCCACT AGAGTTCCCA GCGCAACCAA GGTATCCATG 8400

TTGGCATTGT GCTTTTTAAA ACTGGCCCAA GCACTCTGGA TATATGGCTT ACCTGCAACT 8460

AACATAATAG GCGTTGTTGC TAGAAAGGTT CCCCAATGCA TGACTTGATG ACTAATGCTA 8520

CCTGTCAACA TCCCAATCAT GAGAATCACA AGAGGCACAG TAAAGATACT AGTAATCCAA 8580

AAACGTTGCA GGAGAGATAG AGATTTTCGA GTCTTCTCAA CGACTGTATA GCTTCCCTTT 8640

TGCATCTTCA TGCCACAAGA AAATTCATGT CGCCCTAATT CTTGAGGCGT AAAACGAATG 8700

ACTTTCTCCT CATCTACGCC GATTGGTTCC AAGATACCTT CTTCTTCAAA CAGAATTTCC 8760

TTATAACAGT TTGAAGGAGT AGCACGATGA AAGGTAATCT CAGCTGGAAT TCCCTTTTGA 8820

AGCTGGATAT GGGCTGGATG ATAGCCTTTT TCAGCTCGGA TACGGATTTT TTGAATGCCA 8880

TTTTCTAAGC TTGCTTTCAC AATTTCTGTC ATAGTCTCCA CCTACTCTAC AATCATCTTG 8940

CCGTGCATCA TGTTCATACC ACAAGCAAAG CCAAACTCTC CAGCCTGTTC AGGCGTGATT 9000

TCCACTACAT ACTCTTCCCC CATTGGCAGG TTCGCATGTA CACCAAAATC TGGAAAAACA 9060

ATTTGATCCA GACATGGTGA AGGATCCTTG CGGTCAAAGA CAATGCGTGC TGGCACTGAT 9120

TTCTTGAGGA CAATCAACTC AGGAGTATAG CCTCCCATGA CTTCCACTCG AATCTCTTGG 9180

TATCCGTTTT TTTGCTGGGC TTTTTGTCCA GATTTTTCAG GCTTTTTGAA AAACCAAAAC 9240

AAGATAAACG CGATAAGGGC AATACAAATA ATGGTTACAA TACTATTTAA CATGACGTCT 9300

CCTTTACATA CAATTACATC TTACTTCTGT TACAGCACTT GATTTCTTCT CTGAAATCAC 9360

AGCTTCCAAG TCTTCCAAGT CAGTCTGAGT AAATTCACAT TCTACAATCA AGTCAGCCAA 9420

CAAATTCCTA ATCCTACGGG AACAAACCTT GTCTTTGATA TCTTGGACAA GTAAATCCCG 9480

ACTTTGGTCT AGAGTTAAAA GGGCTGAATA AACAAAGGAC TTGCCTTCTT TTTTCCGAGT 9540

CAAACACTCT TTATCAACCA GACGAGCCAA AAGTGTCTGA ACCGTGGACT TGGACCAGTC 9600

AAACCGCTCT GCCAAAACCC TAATCAAATC TGTACTGGTC TGCTCCCCCT GCATCCAAAT 9660

AATCTTCATG ACCTGCCATT CTGCATCTGA AATCTGCATT ACCATACCTC CAAAATCTAC 9720

ATTTGTCAAT TACACTCATC AGTATACTCT TAAAATCTAC ATTTGTCAAT TATAGAAATA 9780

ATATTTTCTT CGAAAAATAG AATTTTAATC ATTTGAAAAA CGATTTGCAG TCAAATATTA 9840

CTATATAAAC AATAAAAATA TGCTATACTA AAGAAAAAAG AAAACAACCA CTAGGGGTGC 9900

GTAAAGCTGA GATTAACGAC TGTTAGATCC CTCTGACTCA ATCTAGGTAA TGCTAGCTGA 9960

TGGAAGTGGA AATGATAATG GGGACTAGCA GTCTTCTATT GCCTTTCTAA AACAGACTAG 10020

CTTGTTCTTA AGAATACAAA CTTCAGTTGG TTGGGAGGTT TTAGATGACT TATTTACCCG 10080

TTGCTTTGAC CATTGCAGGG ACTGACCCTA GTGGTGGTGC TGGCATTATG GCAGATTTAA 10140

AGTCATTCCA AGCGAGAGAT GTCTATGGAA TGGCTGTTGT AACCAGTCTT GTCGCTCAAA 10200

ATACCAGAGG TGTTCAGCTA ATCGAGCACG TTTCTCCTCA AATGTTGAAA GCCCAATTGG 10260

AGAGTGTCTT TTCTGATATT CCACCTCAGG CTGTAAAAAC TGGAATGTTG GCTACTACTG 10320

AAATCATGGA AATCATCCAA CCCTATCTTA AAAAACTGGA TTGTCCCTAT GTCCTTGATC 10380

CTGTTATGGT TGCTACAAGT GGAGATGCCT TGATTGACTC AAATGCTAGA GACTATCTCA 10440

AAACAAACTT ACTACCTCTA GCAACTATTA TTACGCCAAA TCTTCCTGAA GCAGAAGAGA 10500

TTGTTGGTTT TTCAATCCAT GACCCCGAAG ACATGCAGCG TGCTGGTCGC CTGATTTTAA 10560

AAGAATTTGG TCCTCAGTCT GTGGTTATCA AAGGCGGACA TCTCAAAGGT GGTGCTAAAG 10620

ATTTCCTCTT TACCAAGAAT GAACAATTTG TCTGGGAAAG CCCACGAATT CAAACCTGTC 10680

ACACCCATGG TACTGGATGT ACCTTTGCTG CAGTGATTAC TGCTGAACTA GCCAAGGGCA 10740

AGAGTCTTTA CCAGGCAGTT GATAAGGCCA AGGCCTTTAT CACAAAAGCT ATTCAAGATG 10800

CCCCTCAACT CGGTCATGGT TCTGGTCCAG TCAACCATAC AACTTTTAAA GATTAAGAAA 10860

AAAAACTCTC TAGTTCCCAC TTTAAGGGAA TTAGAGAGTT TTTATACTCT TCGAAAATCT 10920

CTTCAAACTA CGTCAGCTTC CATCTGCAGC CTCAAAACAC TGTTTTGAGC TGACTTCGTC 10980

AGTCTTATCT AAAACCTCAA GGCAGTACTT TGAGCAACCT GCGACTAGCT TTCTAGTTTA 11040

CTCTTTGATT TTCATTGAGT ATTAATTAGG AAAGAATGTT ATGCAACTTT TTTAAAAAGG 11100

CTTGCGTTTT TGCCTCAATA TCTTCTGCTT GCATCAAATC ACGTACAACA GCTACACCAG 11160

CTATGCCAGT GCCCATAAGC TGATCAATAT TCTCCGAAGT CAAGCCTCCA ATAGCAACTA 11220

CTGGAATGGC AACCGTTTGG CAAATTGTTT TCAAGGTCGA TATCAGAGTA ATGGGCGCAT 11280

TTTCCTTGGT GGTGGTTGGG AAAATGGCTC CTGTACCCAA GTAATCTGCA CCTGATTTCT 11340

CCGCTTCCAG AGCTCTTTTA ACCGTTTTAG CGGTGACACC GAGGATTTTT TCAGGACCCA 11400

AGACTTTGCG AGCTACCGAA ACTGGTAATT CATCATCTCC GATATGCAGA CCTGCTGCAT 11460

CAACCGCAAG ACAAACATCC AACCGATCAT CGATTATCAA GGGTACCTGA TAAGCATCTG 11520

TTATTTCCTT GACTTGTTTT GCCAGTTGAT AATATTGATT GGTTGTGAGA TTTTTTTCTC 11580

GCAATTGGAC TATGGTAACC CCTGAACGGC AGGCCGTCTC AACTTTTGCA AGAAAGCTTT 11640

CCACGGAATC TTGATAGCGA TTGGTTACCA GATATAGTCT AAGTGCTTCT CTATTCATAA 11700

ACCTCTCCTT TGATGGTATC TAGCCAATTT TCATCTCTTC TTAGGAGCGA AAGCTGATTG 11760

AGTACTTGGT AACGAAATTC TTCCAATCCC ATTCCTTGAA CAACTATTTT CTCAGCAGCG 11820

ATATTGAGAT AAGAGACTGC TAAGCAAGAA GCTTCAAAAC CAGTCTTTCC TTGGCTGAGA 11880

AAAACAGCTG TTAAGGCTCC AACCAAGTCT CCTGTCCCTG TTATCCAGTC TAATTCAGTA 11940

CAGCCATTTC CCAGTACAGC GACCTGATTT TTCGAAACGA CGAGGTCCTT GGGACCTGTG 12000

ACTAAGAAAG ACATACCAGG ATAGGTCTGA CACCAGTCTT TCAAGACTTG AAGCAAATCC 12060

TCCGTTTCTT GATCTTTAGC ACTCGCATCG ACCCCAACGC CGTGGTGCTT TAATCCAACA 12120

AGACTTCGAA TTTCTGACAT GTTTCCTTTA AGGACCGTAG GTCTATAGTC TAAAAGGTCT 12180

TTAACTAAGC TCTTACGAAT GGATGAAGTC GTTACGCCAA CCGCATCTAC TACCATCGGG 12240

AGAGAAGATT GGTTTGCATA CGAAGCTGCC ATGCGGATTG CTTTTTCCTT CTCAGCTGAC 12300

AAATGCCCCA AATTGATGAA GAGAGCCTGA CTTTGCTTAG TAAAATCAAG AACTTCACGG 12360

GAATCATCTG CCATGACAGG TTTGCATCCC AGAGCCAAAA TCCCATTTGC CAGCATCTCA 12420

CAAGAAATCT CATTGGTAAT GCAGTGAATG AGGGAACTAG AGCCTATAGG AAAGGGATTT 12480

GTAAATTCCT GCATCAGTCT ATCCTTTCAC TAAAGAAATA TCCCTGCACT TTTTTAAAGA 12540

ATTCCTGCTT GATTAAAAAT CGAAAGGCAA TAAAGGAAAT CGCTGTACCA ATCAAGGTTG 12600

CTCCGAAAAA TCGAGGCGTG TAGATAAACC AGCTAAGCTT AGCAGCTGAT CCTGTAAAGA 12660

GTACCATAAC AGGATAGGAA ACAATGGAAC CAATAATACC TGTTCCCAAA ATCTCTCCTA 12720

GAGCAGAATA GTGAAATTTT CGACCGTACT TATAAAAGAG ACCTGCTAGA AGGGCTCCAA 12780

AAGTCGCTCC TGTGAGAGCT AAAGGCGGAA TCCCTTGAGT CGTCATACGG ATAAAGGCTG 12840

TGACTGTAGC CATAGCCAAG GCATAAACAG GTCCCATCAT GATTCCTGCT AGAATATTGA 12900

CTACACTGGA CATCGGTGCC ATTCCCTCAA TTCGAAAGAT AGGTGTAAGG ACTACATCAA 12960

GGGCAATCAT CATAGATAAA ATGGTTAATT TGTGAACTTG TAATTGGTGC TTTCTCATGC 13020

TTCTATTCTT CTCCTTTTTC TAAAGACTGT AAATCGCTCT TCCATGTCTG GTGTTGGTAG 13080

GCCATTTCCC AAAACTTGGC TTCCATATGA ACACTGATGT GGAAGGCATC TAGCATTTTT 13140

TGCTTGTCTG TCTCGTCACT TTCTCGATAG AGCTGATTGA CCAGTGCTCC CTCCTCTCTG 13200

ATCTGTTGCT CTAACTCATC CGTAATATAA GTTTCAATCC ATTGTTGATA GAGAGGATTT 13260

GGTGATGGTT TAAGATTAAG TGATTTGCCT ATATCATGGT ATAACCAAGG ACAAGGAAGC 13320

AAGCTTGCAA AAGCGATGGC TAAGTTCGGT TCTGCAAATT GCCTATAAAT ATGAGAAATG 13380

TAATGATAAC AGGTTGGAGC GATTGGATGT TGCTCCATTT CCTGGTCGCT GATTTCCAAT 13440

TCCTTGAAAA ATTGTTGGCG AATAAATAAC TCACCCTCCA CTAAACCCTG AGCATTTTGT 13500

TTCAAGAGTC TTTTCATCTC TTGGTTTGAA GTCTTATCAG CCAAAAGATG ATAGATTTCT 13560

GAGAAAGCCT TCAGATAGTA GGCATCCTGA ATCAGGTAAT AGCGGAAAAT GGCAGGTTCT 13620

AAATTCCCCT CTTGTAATTG TAAAATAAAG GGATGATGAA AGGAAGCCTG CCAAGCTTTC 13680

TTGGATAATT CCATCGCAAT ATCTGTAAAT TCCATAATAA CTCCTTTATA AAAATAGACT 13740

GGTTTGAAGC AATAAAAAGA AAAGCAGGTA GATTAATTTT GTTTTTTTAG GAATATAAAA 13800

AGTCCGATAG CTATTCTTCA ACTGTGCATG TTCGTCATAT CCGTGAGCAG ATAGAGCTCT 13860

CAGGTAAAGA TGGCGCCACC TAAAGACTGT CATCAGAACC TTACTGTAAA TCAAGGGCGA 13920

CCAAAAATGT AGTTCTTGAC CACGTAATAG GCAAGCTTCT TTGAGGGACT TGATTTCTTG 13980

CTGAATGAGA GGAAAAGAAT TGAATACCAC AATCAAGGCA TAGGACCAAG AGCGTGATAG 14040

CCCCTTTTGA GCCAAGTACA AGAGAAGCTC TTTTAGTGAA ACAGAGGAAA CAAAGACAAG 14100

GCCGATACAA ACTGTCACAA AGGCCCTCGT TCCAAGCATG ACTGCCTGTG AAGCATCTCC 14160

GTGTAACTGA ACTGCCCAGT AGTTGGCAAA AGATGGTAAA ATGGCAAGTA TGATCATCCA 14220

AGCTAACATT TTAAATCGAC GGTAATAGAG CATAAAGAGA ATACAAAATG CGACTACCGA 14280

AAGAGTCAGA GCAATCGAAG GAATGAAAGA TGTTTCCAAG GATAAAATCA GCAAGAAGAG 14340

ACTGATAATC GGTGTCTGGG TTGCTACTTT GACCATACTA TCTCACCTCC CCTTGGGTAT 14400

TGCTACTCTG AGATGTAAGT GGTTTGGTAA TGGTCACTTC TTTCACATGC CGAAGACCCT 14460

GACTAGTCAT CTCAATCCAA TAATCAACCA CAGAAATCAA AGGGTCTAAA CGATGACTAA 14520

TGAGCAGAAA ACTTCTTCCT TGATTCCTCT CCTCCACAAT CCACTTGCAA AAATAATGGC 14580

AGGCTCTATC ATCCAAACCT GCAAAAGGTT CATCTAGCAA GATCACGGAA GCCTTACTGG 14640

TCAAGATGGT CAGGAGCTGA AGAATTTTTT GCTGACCACC ACTTAATTGA TAGGGACTCT 14700

TATCGACTGC CTGCTCCAAA TCAAAATATC GTAAAGCTTG AAAAATCCGC TGATTTCTTT 14760

CAGAATCAGG TCCATCTAAT TGAAGCTCCT CTCGCAGACT GACTCGGATA AACTGCTTCT 14820

CAGCTTCCTG AACAACACCA GTCAGATCAC GATACAAACT CTTTTTCTTT TTCAGGACCG 14880

AACCCTTCCA AGTAATGCTC CCCTTATACT TTTGAAATTG AAGAATAGAC CGAAAGAGGG 14940

TTGATTTCCC GACACCATTG TCACCCAGGA TACAGGAAAT CCCTTGATAG AATGTGAAAT 15000

CAGCAATTGA AAAGAGGGGG CGATTACCAA GCTCACCAGT CACACGGTTC ATATGGAATA 15060

GTTCCGGGCT AGAAGCAACT TCCTTTGAAG CAACCTGTGT CATCTCATAG GAAGGGATTT 15120

GAAACACTTC CCTTAGTTTT CCGTCTCTTA GCTCCACCAT ATGGTCGATA TAGGCTTTAT 15180

AGTCAGATAA ATCATGGTCG CACAAAATAA CTGTCTTCCC ATCATAGACC AACTCTTTTA 15240

GAATCTCCAA TATCTCGATT CTGCTCTTGC GGTCAATGGA AGCGAAGGGC TCATCCAAGA 15300

GATAGACCCT AGGATTCATG GCAAAGAGGA CAGCCAGCGC TGCTTTTTGC TTTTCCCCAC 15360

CTGATAAGTG ATGGATGAGA CGGTGCAAGA TGTCCTTGCA ACGACATTGC TGGACAACCT 15420

CTGCTATTTT AGAATCAATT TCCTGAAGGT GATAGCCGAT ATTTTCCATG GTAAAAACCA 15480

ACTCCTCAAA CAAGCTCTCC ATGGTAAATT GATGATTAGG ATTTTGCAAG AGAATACCAA 15540

CCGTCTGGAC ACGTTCGACG ATAGAAAGCT GACTGACCTC GCTCCCATCT ATCAGGACTT 15600

GACCGCTATA GGGAAGAGAA CTAACTTGGG CAATCATTTG AAAGAGGCTG GATTTTCCAG 15660

ACCCACTACT CCCAACTAAC AAGGTAAAGG CTTGCGCATG AAAAGTAAAA TCAAACGGCT 15720

CAGAGAAGAT TGGGGACTGA ATCGCTCGTA GTTCCAGACC CATCTATGCT TTTCCTCCAG 15780

TTGCAAACTG ATGATAGAGT TTGACAATGG CACGAACCAA GATGGTACAG AAGAAATAAA 15840

CAGAAATAAA ACGTACCACA AGCAAGGAAA GGACAAACGG AAGGGAAAAG GCGTAGTAAC 15900

CTAACTTAAT GTATTCATAG ACAAAGCTAA CAAGCGTAAT CCCAATACTA TTAGCAGTTA 15960

GAGAGAGCCA ACTTTCATAG CGATTCTTAG TTACGATAAA ACCAAATTCA CTTCCCAAAC 16020

CTTGAACAAA GCCAGACAAA AGAGCTCCTA GACCAAATTG GCTACCATAA AGGACTTCAG 16080

CAAGCGCAGC TAGCACTTCT CCAATCGTTG CACTTCCGAC TCTCGGAACA AAGATGGCAG 16140

CAATGGGCGC AGCCATACAC CAGAGACCGA AGAGGATTTC ATTGGCAAAG GCCTGCAAAC 16200

CAAGAGGTGT TAAGAGTAGA CTGAGAATAT TATACACATA TCCTGAACCA ACGAAAACCC 16260

CACCAAAAAA GATAGACAAG AAAGCAAGCA AGATAACATC TTTTAACTGC CATTTTTTCA 16320

ACATAAAAAA CTCCTTTTTT TAAAGAAAAG TGAGGCACTC AAGAAGACCG ACCTAAATAC 16380

TTTGTATAGC AGACTGAATT TAGAACAGTA CACAAGAACA CTAAAATATT TCTAGAAATT 16440

AATTTGAATT TTCTAATTGA TTTGTTCGCA TCTTATTTCA ATCTACTATA TCATCTTCAT 16500

CCAGTTTCGT AAAAGAAAAA ACTCTAATTA CAGATACAAA TTAGAGTTCA GCTTACAAGA 16560

TTAGACAGTT CTTTTCGACA TACGAAAAAA ACATTTCACA TTTCCCTTCG CCAGTCTTAA 16620

CTGTATCAGG TTCAATGGGT ATCATCTCAG CCTAAAGCAC CCCAAATGTC TTTATTATTT 16680

AATTATGTGA TTATTATAAC ACACATTTTA TACTAGTTCA AGAAATTGAA CTGGAAATAC 16740

AGCCTTGCAC TCACAAAGAC AGCAGATCTT TCTTTTGCAA AAAACAAATG ACCTGTTTGA 16800

TGAATTAGCC ATTCAAGCTG AATCTGGACA TAGCTTTTTA AAAAAGGAAA ATCCTACTTA 16860

CTTAGAATCC AAGGATAGAT ATCTATTGTT CACTCATTTC CCGAACAGTT TTTTCTATAT 16920

TTTTTGCATA CGATATTGCC GAAATGATTG AAACGCCATC CATATTGGTC TTTATAATGT 16980

CTTTAATATG TTTCGTCTGT ATCCCACCAA TTGCAACTAA AGGCATTTGT GGCAATAGTT 17040

TTCTCATCAA TTCAAGACCT TCATAACCTA TAGTACCACC AGCATCATCC TTTGACTGGG 17100

TACCAAATAC AGGCCCAACA CCTACATAAT CTACATATTC AACTTTTGAT TGTTGAAATT 17160

CTTCTTCGTT TCTTATAGAA AGACCAATTA TTTTATCTGG CATCAATTTT CTAATTTCAT 17220

CAACACCAAT ATCATCTTGA CCTACATGTA CGCCATCGGC GTCAATTTCC ATTGCTAAAT 17280

CTATATCGTC ATTAACGATA AATGGAACAT TGTATTTTTT ACAAAGTTCT TTAATTTGGA 17340

TAGCTAGCTC AAGTTTTTCT AAGCCTTCTA AAGCACCCTC ACCTTTTTCT CGAAATTGAA 17400

ATAAGGTTAT ACCACCTTTT AAGGCTTCCT CAACGACTGT ATATAGATTT TTTCCTTGGC 17460

AAGTAGTCGT TCCACAAATA AAATATAGTT TTAGTAATTC TTTATGAAAC ATCTTACTTC 17520

ACTCTTTTGA ATTCCTTTAC ATCTTCATCT GTAATCTCGT ATAAGGCATT TATAAATTCA 17580

ACTTTAAATG TCCCAGGAAG ATGTCCATTT GGACGTTTTT CTGCTATTTC TCCAGCGATA 17640

TTGTAAACCA ACACTGCTGT TTTTAATGAT TTCAATTCTT GACCTTTTTC TAGTCCGATA 17700

AAGCTTGCTA CTACAGCTCC TAATAAGCAT CCTGTCCCAA TGACTTTCGG CATCATAGCA 17760

CTACCATTAT GAATCATTAC CACTTCTCCA TTAACAGCAA TGGCATCCAC TTCACCTGTT 17820

ACTACTATTG GAATATTGAA CTTCTCATTT GCTGCTAGAG CAATTTCGTC AATATTATCT 17880

ACGCCCGCAC TATCTACTCC TTTAGATGCC ACATCTATTC CTACTAAAGA GGCAATCTCG 17940

CCAGCATTTC CTCTAATCGC TGCTAGTTTA TAATTGTTGA TTAGATCATC TGCTACTTTT 18000

TTTCTATATT CTCCTGCTCC ACAGGCTACA GGATCTAAAA CTGCTGGGAC ATTATATTTC 18060

TCTGCAATTT TCAGAGCAGC TTGGTATAAT TTCCAATTTT CATCTGTCAA TGTTCCTATG 18120

TTTATTAATA AACCACCAGC ATACTTTAAC AAATCCTCTA AATCTGCTGG AAACTCACTC 18180

ATGGCTGGTG AGGCGCCCAG TGCTACTAAT CCATTTGCTG TGAAATTTTT TACTACATCA 18240

TTGGTTATAC AAATGACCAA TGGTGCTTTT TCTTTTAATA ATTTTAAACT TGTCATATTG 18300

AAATCCTTCC TTTTCACTTT ATACGATCTA CTAATTTCGA TTTATCTTTA GTTGAGAATT 18360

TTTTTCATTT ACATTGAATG ATTTATACTC AATGAAAATC AAAGAGCAAA CTAGGAGGCT 18420

AACCGCAGGT TGCTCAAAAC ACTGTTTTGA GGTTGTGGAT AGAACTGACG TGGTTTGAAG 18480

AGATTTTCGA AGAGTCTTAC CTCATCAAAT TTGTAAATAT CATGAGCCTT CTCTAGACAT 18540

CGTAACCAAT ATCAAAAAAA GCTAATTCTA AAGCGACTGC TTGATTCCAG CGTTGCTGAA 18600

GTTCTGTCAA ATCTTCTCGA TTTTTACCGA CACGATTGAG TTCGTCAACC AGAAATTGAA 18660

CCCACTCTGC AAAGAAAGGA CCTCTGTGGA GATTGATCCA TTCCGAATGA ATATAGACTT 18720

CAGGTAAAGC CAAATCTTTA GAACCCCAGT CTAAATAGAG ACCTTCTGCA ATGACCAGCA 18780

TGACCAAAAG ATGGGCATAG TCTGATGAAG CCACCGCCGA ATACATTAGA TCCTGAAAGG 18840

CTTTTGTTAC AGGGTGCAAA GTCACTTCTA GATAGTCATT CTCTGCTACT TTTAACTCTT 18900

TAAAAGCCTT TTGGAAATAA CCATCTTCAT CTGCTTCAAG AAAGCCTAGT TGCTTGGCAA 18960

AACGAAGCTT GGATTCAAGT TTATCTGCGT GACTACGCAG GCACCCAGCA TGGATAAGAA 19020

GGCATCAAAG AAGTGATAAT CTTGAATCAG ATAGTCCTTT AAGACCTTAT TCTCAATTGT 19080

CCCCGCAAAA AGTTCCTTAA CAAAACGATG ATTGATTGCA GCCTGCCAAT CCTTCTGACT 19140

GCTTTTTAAT AATTCTCCAA CAGTCAAACC TGGCTGAAAT GCATAGTCTT GTGTTTCCAT 19200

ATTTACTTCT CCTCTCTTTA CTTGTTAGTA ATTAATAAAA CACCAAGAAA TATCAAGCAA 19260

AATCGTAATT CCACTTGATC CTTTTAAAGC ACATCGAGAG CATT[illegible]AGA GAGCTAACTA 19320

AACAAGCCTA TCCAGTTTAT ATAAACAAAA AACTCCAATT ACAATCAAGA ATTAGAGTTG 19380

ACTTACAAGA TTAGACCGTT CATTTCACCA TACGAAAAAA CTGTTCACAT TTCCCTTCGC 19440

CAGTCTTAAC TGTATCAGGT TCAATGGGTA TTATCTCAGC CTAAAGCACC CCAAATGTCT 19500

CTATTATTTA ACTACTGAAC CAGTATAGCA AAAAATGAAA GCCCTAGCAA GATATTTGAC 19560

CGAAAAATAT CTTTATATAT AATATATTGA AACTAGAATA GTACACCTCT ACTTATAAAA 19620

CATTGTTAGA AATCGATTTG ACTGTCCTGA TTGATTTGTC CTATTCTTAT TTCATTTTAC 19680

TATAGTTTTC GATAGCAATT TATTCTTCCA ATACACGAAG AAAAACCTCC ACATTCAGTG 19740

GAGGCAATCT GTTTTATCAA TACAATTTTA AGTCACGAGG GTCAACTGGG AAGGTTGGGT 19800

TGTATGGATT GTGACGGAGC TTGAAGTGTT TGACATCTTC AATGGTCTGA GTTCCAGACA 19860

ATTGCATAAC TGTCTTCAAT TCCGCATTCA AGTGTTCAAA GACTTGACGC ACACCGACAC 19920

TACCACCGAG AGCCAAGCCA TAGATGACAG GGCGTCCAAT AGCAACCAAG TCTGCTCCTG 19980

ATGCCAAGGC TTTAAAGACG TGTTGACCAC GACGAACACC AGAGTCAAAG ACAATCGGCA 20040

CACGTCTATC AACTGCTTCT GCCACTTCTT GAAGCGAGTC AAAGGCAGCT GGTCCACCGT 20100

CGATTTGACG ACCACCGTGG TTGGTTACCC AGATACCAGA AGCTCCTGCA GCAAGCGAAC 20160

GTTCAACGTC CTCACGGCAT TGTGGTCCCT TGACATACAC AGGAAGACCA GAGTATTCAG 20220

CGATAAATTC TACATCGCGT GGAGACAAGC GTTGTTTAGC TGATTTGTAA ACAAAGTCCA 20280

TTGATTTACC AGCACCTTCT GGCAGGTATT CTTCAACAAT CGGCATGCCA ACTGGGAAGA 20340

CAAAACCATT ACGCTTATCC ACTTCACGAT TCCCCCCTAC AGTAGCATCT GCCGTCAAGA 20400

CAATCGCTTT ATAACCTTCA GCCTTCACAC GGTCCATGAT GTGGCGGTTG ATACCGTCAT 20460

CCTTACTAAA GTAAAATTGA AACCAATGAG GTGTCCCTTG GAGGGCTTCA GAAATCTCTG 20520

GAAGGTCAAC AGTAGAGTAA GAACTGGTTG TATAAAGAGA ACCAAACTCA TGCACACCAC 20580

GCGCAGTCGC CACTTCCCCC TGTTCATTTG CCAATTTATG AGCCGCAACA GGTGCCATAA 20640

TGATTGGAGA AGATAGTTTT TCACCTGCAA ATTCAATCTC TGTACTTGGA TTTTCTACAT 20700

TGCAAAGTGT ATGAGGAACG ATGAGCTTGT GGTTAAAGGC ACGGATATTC TCTCTTAAAG 20760

TGAAAGTATC TTCCGCCCCA CTAGCGATAT AGCCAAATGC TGCTTTAGGA ATAACTTGTT 20820

GCGCCATTGG CTCCAAATCA TAGGTATTGA TGAArTCTAC ATGACCTTCT GCATTGCTTG 20880

TTTTGTATGA CATAAAATGT CCTCCTTAAT AAGTAAGCGT TTACTTTGTG TATTACAAAA 20940

ATATCTTAAC TCTTTTTCAA AACTTTTAAA ATATTTTGTT TGGAAATTTC AGAAATTTTA 21000

TGTCTATGAT AAAAATCCTT ATAACGGCAA TAAAAAATAG ATATTATCCA AAGAAGATTT 21060

TAAGTGCTAC AATAACTGTA TTATTTCTAG ATGGGAGGTT CTATTTTTGG ATTGATCCAT 21120

TGTTGAACAA TATCTACCAC TATATCAAAA GGCATTCTTT CTGACCTTGC ATATTGCAGT 21180

TTGGGGAATT TTGGGATCCT TTCTGCTCGG TTTAATCGTT AGTATCATCC GACATTATCG 21240

AATCCTTGTT TTGGCGCAAG TAGCGACAGC CTACATTGAA TTGTCACGTA ATACGCCCCT 21300

TTTGATTCAA CTCTTCTTTC TCTACTTCGG TCTTCCCCGA ATCGGGATTG TCCTATCTTC 21360

AGAAGTCTGT GCAACGCTTG GGCTTGTCTT TTTAGGAGGC TCCTATATGG CAGAATCTTT 21420

CCGAAGTGGG CTGGAAGCCA TCAGTCAAAC CCAGCAGGAG ATTGGCCTCG CTATTGGTCT 21480

GACACCTCTA CAGGTCTTTT ACTATGTGGT TCTTCCGCAA GCAACAGCGG TGGCACTCCC 21540

CTCCTTTAGT GCCAATGTCA TTTTCCTTAT CAAGGAAACC TCTGTTTTCT CAGCAGTGGC 21600

TTTGGCCGAC CTCATGTACG TCGCCAAGGA TTTGATTGGT CTCTACTATG AGACAGACAT 21660

TGCGCTAGCT ATGTTGGTAG TTGCTTATCT AATCATGCTG CTACCCATCT CACTGGTCTT 21720

TAGCTGGATA GAAAGGAGGC TCCGCCATGC AGGATTCGGG AATCCAAGTA CTCTTTCAAG 21780

GAAATAATCT CCTGAGAATC TTACAGGGAT TGGGCGTTAC GATTGGGATA TCCATCCTGT 21840

CTGTCCTCTT ATCCATGATG TTCAGAACAG TCATGGGAAT CATCATGACC TCCCATTCTA 21900

GAATCATACG ATTTTTAACA CGATTGTATC TGGAATTTAT CCGTATCATG CCCCAGCTGG 21960

TGCTACTCTT CATCGTTTAC TTTGGCTTGG CTCGAAACTT TAATATCAAT ATCTCAGGTG 22020

AGACTTCAGC TATTATCGTT TTTACCCTCT GGGGAACAGC TGAAATGGGA GACTTGGTAC 22080

GTGGAGCTAT CACTTCTCTC CCTAAACATC AGTTTGAAAG TGGACAGGCA CTCGGCTTGA 22140

CTAATGTTCA ACTTTACTAC CACATCATCA TCCCACAAGT CTTAAGAAGA CTGCTACCGC 22200

AGGCTATCAA TCTTGTCACT CGGATGATTA AAACCACTTC ATTAGTTGTT TTGATTGGGG 22260

TTGTGGAAGT GACCAAAGTT GGACAACAAA TCATCGATAG CAATCGCCTG ACCATCCCAA 22320

CTGCTTCATT TTGGATTTAT GGAACCATTC TAATCTTATA TTTCGCAGTT TGCTACCCTA 22380

TTTCCAAACT ATCCACTCAC TTAGAAAAAC ATTGGAGAAA CTAAATGTCT GAAACTATCT 22440

TAGAAATCAA GGAACTAAAA AAATCCTTCG GAGACAATCC CATCCTCCAA GGACTTTCTC 22500

TAGAAATCAA AAAAGGGGAA GTTGTTGTCA TCCTAGGGCC ATCTGGTTGT GGGAAAAGTA 22560

CCCTCCTTCG TTGCCTCAAC GGCTTAGAAA GTATTCAAGG TGGAGATATT CTTCTGGATG 22620

GTCAGTCTAT CGTTGAAAAT AAAAAAGATT TTCACCTAGT TCGCCAAAAG ATTGGCATGG 22680

TCTTTCAAAG TTATGAACTC TTTCCCCATC TGGATGTCTT ACAAAACCTC ATCCTAGGCC 22740

CTATCAAAGC TCAAGGAAGG GACAAGAAAG AAGTAACGGA AGAAGCTTTG CAATTACTAG 22800

AGCGTGTCGG TTTGCTGGAT AAACAACATA GCTTTGCCCG TCAATTATCT GGTGGACAGA 22860

AGCAACGTGT TGCAATTGTC CGTGCCCTCC TAATGCATCC AGAAATCATC CTTTTTGACG 22920

AGGTGACTGC TTCGCTGGAT CCAGAAATGG TGCGTGAGGT GCTGGAACTT ATCAATGATT 22980

TGGCCCAAGA AGGCCGTACC ATGATTTTAG TAACCCACGA AATGCAGTTT GCCCAAGCCA 23040

TTACTGACCG GATTATCTTC CTCGACCAAG GGAAAATCGC TGAAGAAGGA ACAGCTCAAG 23100

CCTTCTTTAC CAATCCGCAA ACCAAACGAG CCCAGGAATT TTTAAACGTC TTTGACTTTA 23160

GCCAATTCGG CTCATATCTA TAAAGGAGAT TCTTATGAAA CTATTCAAAC CACTCTTAAC 23220

TGTTTTAGCA CTTGCCTTTG CCCTTATCTT TATCACTGCT TGTAGCTCAG GTGGAAACGC 23280

TGGTTCATCC TCTGGAAAAA CAACTGCCAA AGCTCGCACT ATCGATGAAA TCAAAAAAAG 23340

CGGTGAACTG CGAATCGCCG TGTTTGGAGA TAAAAAACCG TTTGGCTACG TTGACAATGA 23400

TGGTTCTTAC CAAGGCTACG CTACGATATT GAACTAGGGA ACCAACTAGC TCAAGACCTT 23460

GGTGTCAAGG TTAAATACAT TTCAGTCGAT GCTGCCAACC GTGCGGAATA CTTGATTTCA 23520

AACAAGGTAG ATATTACTCT TGCTAACTTT ACAGTAACTG ACGAACGTAA GAAACAAGTT 23580

GATTTTGCCC TTCCATATAT GAAAGTTTCT CTGGGTGTCG TATCACCTAA GACTGGTCTC 23640

ATTACAGACG TCAAACAACT TGAAGGTAAA ACCTTAATTG TCACAAAAGG AACGACTGCT 23700

GAGACTTATT TTGAAAAGAA TCATCCAGAA ATCAAACTCC AAAAATACGA CCAATACAGT 23760

GACTCTTACC AAGCTCTTCT TGACGGACGT GGAGATGCCT TTTCAACTGA CAATACGGAA 23820

GTTCTAGCTT GGGCGCTTGA AAATAAAGGA TTTGAAGTAG GAATTACTTC CCTCGGTGAT 23880

CCCGATACCA TTGCGGCAGC AGTTCAAAAA GGCAACCAAG AATTGCTAGA CTTCATCAAT 23940

AAAGATATTG AAAAATTAGG CAAGGAAAAC TTCTTCCACA AGGCCTATGA AAAGACACTT 24000

CACCCAACCT ACGGTGACGC TGCTAAAGCA GATGACCTGG TTGTTGAAGG TGGAAAAGTT 24060

GATTAGTCAT TAACTCTTAA AAGGAACTGG ATTTTAAGCT CCAATCCCTT TTTAAGATTT 24120

TACCTATAAC ATCCTGAGTC TATCTAAGAT GTTCAATCTG AACACAGTGT ACATACTTTA 24180

TCTTCTATTG CATATACTTT ATCACATAAG ATACGAATAT CCTCTTCACT ATGACTAGCA 24240

ATCAAAATTG TTGTCCCTTT TTCACTAGAG AGCTTTCTAA ACAATGTTCT CATATTTTCT 24300

ACACTTGATT TATCCAAGGC ATTCATAGGT TCATCTAGTA AAAGAATAGA GGGATTCTCC 24360

ATAATTGCTT GAGCAATCCC TAGCTTTTTC CTCATACCTA GCGAATAAGT TTTAACTTTC 24420

TGGTCTTTTT GCTCATATAG ACCAACTATT TTCAGTGTAT CATTGATTTC CTGATTACCA 24480

ACTACTCCTC GTATGCTTGC CAAATATTGT AAATTCTTAA AGCCACTATA ATAATTTATA 24540

AAACCAGGTT CTTCAATCAA AGCTCCCAAA TTAGCTGGAA TTTTTCTCTC AGGAACAATA 24600

TTTTCCCCAT TGATTAACAC TTCTCCATAA GACGGACTAT ATAAACCAGC TATTAATTTA 24660

AACAATACAC TTTTCCCTGA GCCATTCGCA CCAGTAATTC CTATAATTTC CCCCTGTTTA 24720

CAACTAAAGT TAAGGTTTTG AAAAACACAT GTCTTTTTTA ATTTCAACTC AATATTTTTT 24780

AATGTAATTA TTTCATTCAT TCTATAAACC TCCTCTTTTG ACGAGTGAAA TAGAAAATGC 24840

TTTGAAAAAG AAAGACTAAA AATAGCAACT GAAGAAATAA ATCTCGTCCT ATATCTCCAT 24900

TCCCTCGATT CAAAATATAA AATAGATAAT TAGTTCGATT TCCTACAAAT AGACCACCAA 24960

ACACAATCAT GAGTAAAAAG AAACTAACGC AAGCAAAGTT CG 25002

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11443 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

CAGGTACGGT GAGGCGCAAC TAAAATATAA TTTTCATCTT GATTAGGAAT TTTATCAGTA 60

TTATGATAGT GAGCATTGCC ATTGATGGAC CATAAGAGCA ATACAACTAA TCCACGCAAA 120

TAAGTATAAA ACATGCGATC TCCTTCGATT GTTTTCTTGT TATTATTATA CCTTATCAAA 180

GGAGGGCTGG CAAACTTTTC CCTTGACTAG ATACATATTT AGGATGAAAT TAGAATTCTG 240

TTAAAAAAAA TGATATAATA GAATTTATGG ATAAAAATAA GATTATGGGA TTAACCCAAA 300

GAGAAGTCAA GGAAAGACAG GCTGAGGGTT TGGTCAATGA CTTTACCGCA TCAGCCAGTA 360

CCAGCACTTG GCAAATCGTT AAACGAAATG TCTTTACCCT TTTTAACGCT TTGAACTTTG 420

CCATTGCTTT GGCTCTTGCC TTTGTGCAGG CTTGGAGCAA TCTGGTCTTC TTTGCTGTTA 480

TCTGCTTTAA CGCTTTTTCT GGGATTGTGA CCGAGCTACG AGCCAAACAC ATGGTGGACA 540

AGCTCAATCT CATGACCAAG GAAAAGGTCA AAACCATCCG TGATGGTCAG GAAGTTGCTC 600

TTAATCCTGA AGAATTAGTG CTAGGAGATG TCATTCGTTT GTCTGCAGGA GAGCAGATTC 660

CTAGTGATGC CTTGGTTTTG GAAGGCTTTG CGGAAGTCAA TGAAGCCATG TTAACGGGAG 720

AAAGTGATTT GGTGCAAAAG GAAGTTGACG GCTTACTTTT GTCAGGAAGT TTCCTAGCCA 780

GTGGGTCAGT TTTATCTCAA GTTCACCATG TCGGTGCAGA CAACTATGCT GCCAAACTCA 840

TGCTTGAGGC TAAGACCGTT AAACCCATCA ACTCCCGTAT CATGAAATCG CTGGACAAGT 900

TGGCTGGTTT TACTGGGAAG ATTATCATTC CCTTTGGTCT GGCTCTCTTG CTGGAAGCCT 960

TGCTTTTAAA AGGCCTGCCT CTCAAGTCAT CCGTTGTAAA CTCGTCGACA GCTCTTTTGG 1020

GAATGTTGCC TAAGGGAATT GCCCTTTTGA CCATTACTTC GCTCTTGACT GCAGTGATTA 1080

AGTTGGGCTT GAAAAAGGTC TTGGTGCAGG AGATGTACTC TGTTGAGACC TTGGCGCGCG 1140

TGGATATGCT CTGTCTGGAC AAGACGGGTA CCATCACCCA AGGAAAGATG CAGGTGGAGG 1200

CTGTTCTTCC GTTGACGGAA ACGTATGGTG AAGAGGCTAT TGCCAGCATC TTGACTAGCT 1260

ACATGGCCCA TAGTGAGGAT AAGAATCCAA CTGCCCAAGC CATTCGCCAG CGTTTTGTGG 1320

GAGATGTTGC TTATCCTATG ATTTCCAATC TTCCCTTCTC GAGCGACCGC AAGTGGGGGG 1380

CTATGGAGTT AGAAGGCTTG GGGACAGTTT TCTTAGGGGC ACCTGAGATG TTGCTTGATT 1440

CTGAAGTCCC AGAAGCTAGG CAGGCCTTGG AGAGAGGATC ACGTGTCTTG GTCTTAGCTC 1500

TCAGTCAGGA GAAATTAGAC CATCACAAAC CACAGAAACC ATCTGATATT CAGGCTCTAG 1560

CCTTGCTGGA AATCTTGGAC CCCATTCGAG AGGGAGCAGC AGAGACGCTG GACTATCTCC 1620

GTTCTCAGGA GGTGGGACTC AAGATTATCT CTGGTGACAA TCCAGTTACG GTGTCCAGCA 1680

TTGCCCAGAA GGCTGGTTTT GCGGACTATC ACAGCTATGT AGATTGCTCA AAAATCACCG 1740

ATGAGGAATT GATGGCCATG GCGGAGGAGA CAGCTATTTT CGGACGTGTT TCCCCTCATC 1800

AAAAGAAACT CATCATCCAA ACGTTGAAAA AAGCGGGACA TACAACGGCT ATGACAGGGG 1860

ACGGGGTTAA TGATATCTTG GCCCTTCGTG AGGCGGATTG TTCTATCGTG ATGGCGGAGG 1920

GGGATCCAGC AACCCGTCAG ATTGCCAATC TGGTTCTCTT GAACTCAGAC TTTAATGATG 1980

TTCCTGAGAT TCTCTTCGAG GGTCGTCGCG TGGTCAATAA CATTGCCCAC ATCGCCCCGA 2040

TTTTCTTGAT AAAGACCATC TATTCCTTCC TGTTAGCAGT CATCTGTATT GCCAGTGCTT 2100

TACTAGGTCG GTCAGAGTGG ATTTTGATTT TCCCCTTCAT TCCGATCCAG ATTACCATGA 2160

TTGACCAGTT TGTGGAAGGT TTCCCACCAT TCGTTCTGAC TTTTGAGCGA AATATCAAAC 2220

CTGTTGAGCA GAATTTCCTC AGAAAATCCA TGCTTCGTGC CCTACCAAGC GCTCTCATGG 2280

TCGTCTTCAG CGTCCTGTTT GTGAAAATGT TTGGCGCGAG TCAAGGTTGG TCTGAGTTAG 2340

AAATCTCAAC TCTACTCTAT TATCTCTTGG GGTCAATTGG TTTCTTATCC GTATTTAGAG 2400

CCTGCATGCC ATTTACCCTA TGGCGTGTCC TCTTGATTGT TTGGTCAGTA GGAGGTTTCC 2460

TAGCCACAGC TCTCTTCCCA AGAATTCAAA AACTGCTTGA AATTTCAACC TTAACAGAAC 2520

AAACGTTGCC TGTTTATGGT GTCATGATGT TGGTCTTTAC CGTGATTTTC ATCCTGACCA 2580

GTCGTTACCA AGCGAAAAAA TAAATCAAAA CCACCAGTGT GAACTGGTGG TTTGTTCTGC 2640

GGCTATAAGC CGCTTCTACC GGCCAGGGCC AAAGGCCCAC CGAAATAGCT TCCTCGCGCA 2700

CCACTTTCCC GAGCAGGTGC TAAAGCACCT TAGTTACTTC CTCTTATTTA TTTCGCCAGT 2760

AAACGGATCT ACTGACTCGA ATAACGTGAG CTGGTCTGCT ACTCTGTCTT CTTGTAATTG 2820

ATTCTGAATA TATTCAGCTA TCACTTTCTG ATTACGGCCT ACCGTATCTA CATAATAGCC 2880

TCTACACCAA AACTTGCGAT TGCCATATTT GTATTTTAAA TTCGCATGCT TATCAAAAAT 2940

CATCAAACTG CTCTTGCCCT TTAAATAGCC CATAAAGGAC GAAACACTAA GTTTCGGAGG 3000

AATACTGATA AGCATGTGAA TATGGTCTGA ACAAGCATTC GCTTCATGGA TTATTACACC 3060

CTTACGCTCA CATAAGTCAC GTATGATTCT TCCGATACTA GCTTTGTATC TGCCATAAAT 3120

GATTTGACGA CGATATTTGG GTGCAAAAAC AATATGATAT TTACAATTCC ATGTGGTATG 3180

TGATAAACTT TGATTATCCT CTCTCATGAG GTACCTCCTG TATGATATGT TGTAGTGGCG 3240

GAGAAACCAC TTCTATCTTA TCATTTTAGG AGGTTCTTTT TGTTACCACG CTAAAAGCTC 3300

TATGGAACCA CTAGCATAGC TAGTGGTTTT CGGGAGACAA CAAGAAAGAC TGCAATCTGT 3360

GGATTGCAGT TTTTTATACG ATGGATCTAT CGTAGATCTG ATGTGCAAGG CCTACGTGCC 3420

GATCATCTAT CGGTGAACCC AAGAGCGACC CTCAAGCCTG CTTGGATTGA GGTAATAGAT 3480

TCAAATATCT GTAGTTAGAC TATTTGAAGT TTGATGTAAG AAAGAGAAAG CGACAGATTG 3540

AAGTAATTTT AACTCTCTTC TATTGCTAGA ACAAATGGTC GGATAGGTTG GTAGTTTGAA 3600

AATGAAGATG CTATCTATTG TTAAATGGAA CATAGTGTTA TTTATTAGAA AATCGTTTGG 3660

TTTATTTCTT ATCAAATACG AAAAGCAACT TAAATATTTC AACTAAAATA GATGTTATGA 3720

AGAAAAGGTA AAATGATTTT GGCATAGTGA GGTTCTGTTC TATTTGATAT CATATTTTTG 3780

ATAAAAACAA AAATGTCCAT TGCAAAGGAC AAAATGCGAA GTATATTATT TTTTGAAAGC 3840

GATATAATGG ATTCATAAAG GAGGTGTATC GTGTCTAGAA AACAAGAACA AATGGAAACG 3900

TTGTTGCTCC TTTTGCGAGA TAGTAAGGAT TATATATCTG CTAAAGTATT GGGAGAAAAA 3960

TTAAATTGCT CTGATAAAAC GGTTTATCGC CTTGTCAAGG GAATCAACAA AGATTGTCCG 4020

GTAGAAGCAT TCATTTTATC TGAAAAAGGC AGAGGTTTCA AATTAAATCC AAGAAGTTCC 4080

CTCGTGGACG TTGATGGGAA TTTTACAGAG GCTTTTGATC CTGAAGTAAG GCGTGAAAAA 4140

TTACTAGAAC GTCTCTTGTT GACTGCTCCT AAGCCACATT CTATTTATGA TTTAGGAGAG 4200

GAATTCTACG TAAGCGAGTC AGTAGTACTA AAAGATCGTC AGATATTACA AGAGAGTCTA 4260

GCAATTTATG GGTTAGATTT AAAAATGAGA CAACGAAAGC TTTTTATTGA TGGGGATGAG 4320

GCTCAAATTC GTTCAGCCAT TCTAAATCTA CTGCCAATGT TTAATCAGTT GGATTTAGAG 4380

CAAATTACAC AGAATAAGGT TCAGCCTCTT GACGGAGAAC TTGCTCACTT TTGTTTGGGA 4440

TTACTGATTA CACTTGAGAG AGAATTGGGG GTAAACATTC CCTATCCATA TAATATAAAT 4500

ATTTTCTCTC ACCTGTATAT TTTTATCAGT AGGAATCGTC GTAGTACTAG TATTCATGTT 4560

GTAGCACCTT CAAAACCTAC TATTGTTGAT GAGAAAATTT ACAGTGTCTG TCAAAAAATT 4620

ATTCAAGAAA TTGAACAATA TTTTAGGATG AAGGTTGATG CAGTTGAGAT TGACTATCTT 4680

TATCAATACG TTGTATCTTC GAGATTGCAA AAACCATTTT CTTCCGGGAA GCTTCCTTTT 4740

TCTCAGCGAG TTTTAGATGT CACTCATTAC TATTTTAGCC GTATGTGTAT GGACAATAGA 4800

GAGATTGAAA CGACAGATCC TGACTTTGTT GACTTGGCGA GTCATATCAG TCCCTTACTG 4860

AGGAGATTAG ATAATAGAGT ACAGATTAAG AATAGTCTTT TATCACAAAT TCTTTTAACC 4920

TATCCTAATC TGGTTAAAGA GTTAACAACT ATTTCTAAAG AAGTGAGTCT AGTATTTGGT 4980

TTTGCTTCCT TGAGTCTGGA CGAGATTGGT TTTCTAGTCT TATATTTTGC ACGGTTTCAA 5040

GAAAAGCGAG CACGTCCTCT AAAAACAGTA GTGATGTGTA CATCAGGTGT CGGAACTTCA 5100

GAGCTTTTAC GAGCACGATT AGAAAAGCAA TTTTCTGAAT TGGATATTAT TGATGTAGTT 5160

GCTTATCATC AATTAGATGA GCTGATAAAT CTATATCCAG ATTTAGATTT CATTGTGACG 5220

ACGGTAGCTT TGCAGGAACC AGCAAGTGTC CCGTTTGTCC TAGTTAGTGT TTTTCTAACC 5280

GAGGGTGATA AACAACGTCT TCAAGCAAAA ATTCAGGAGA TAAACTATGA ATAATCTTTC 5340

GCTTGTCCTT ATGGATATAT CTGTTCAAAA TCGTCAAGAA GCCTACAAAG AATTAGCAAA 5400

TCAAATCAGC CTTCTTGTTT CTGAAGATAC AGAAAAAATA GAAGAGCTTC TATATTACCG 5460

TGAGAGACAG GGAAGTATAG AGGTTGCTAA AGGTGTTCTT CTACCACATT GTGAAGGAAA 5520

CTTTCAACAT CATGTCTTAG TGATTACTAG ATTAAAATCA CCTATCAGAG AATGGTCGAA 5580

GGATATCCAG TGTGTTGACC TTATTATCGG TTTGGCCATT GCAGTATCAC AGGACAAGTC 5640

ATGTATTAAA ACATTGATGA GAAGACTAGC AGATGAATCA TTCATAAATC AATTAAAACA 5700

GTTAACAAAA GAAGAATTAC GGGAGATAAT ATATGGAAAT CAAAGATATT CTTAATGTGA 5760

GTCTGATCCA GACGGATTTA CAGATGCAGA GCAAAGAAGA GGTTTTTGAG GCATTAGCTC 5820

AACTATTGGT TGAGACGGGT TATGTGTCTG ATAGAGACCA ATTTATCGAA GGTCTTTATC 5880

AGAGAGAGGC AGAAGGACAG ACCGGTATTG GGAATTATAT TGCTATTCCC CATAGCAAGA 5940

GTTCTGCTGT GGAGAAGGCG GGGGTAGTCA TAGCTATAAA TCACAATGAG ATTCCTTGGG 6000

AGACCATTGA TGGGAAAGGG GTCAAAGTAA TTGTACTCTT TGCAGTTGGT GATGATACAG 6060

AAGCTGCTAG GGAGCATTTG AAGACCTTAT CACTCTTTGC TCGAAAACTT GGTAATGACG 6120

AAGTTGTTGC CAAATTAGTT CGGGCTCAGA CATCTGATGA TGTGATTGCA GCTTTTTGTT 6180

AATAAGAAAA AATTTTGGAG GGTATCCGTA TGAAAATTGT TGGTGTTGCA GCTTGTACTG 6240

TGGGAATTGC CCACACTTAT ATTGCACAGG AAAAATTAGA GAATGCCGCA AAGGTAGCTG 6300

GACATGTGAT TCATGTTGAG ACTCAGGGGA CAATAGGGGT AGAAAATGAA TTGAGTCAAG 6360

AGCAGATTGA TGCAGCGGAT GTAGTTATTT TAGCAGTTGA TGTTAAGATT TCTGGTATGG 6420

AACGCTTTGA GGGTAAAAAG ATTATCAAGG TTCCAACAGA AGTGGCAGTC AAATCTCCCA 6480

ATAAACTGAT TGCTAAAGCT GTTGAGATTG TTACGAAATA ACTGAAAATA TTTAAGGAGA 6540

AAATATATGT TGAAACACTT AAACTTAAAA GGTCACTTAT TGACAGCCAT TTCCTATATG 6600

ATTCCAATTG TTTGTGGTGC AGGATTCTTA GTTGCCATTG GTTTAGCAAT GGGGGGTGGT 6660

GTTCCTGACG CTCTTGTAGC AGGAAAATTC ACTATCTGGG ATGCTTTAGC AACTATGGGT 6720

GGTAAAGCCC TTGGTCTCTT GCCAGTTGTT ATTGCTACAG GTTTGTCTTA CTCGATTGCT 6780

GGTAAGCCAG GGATTGCACC AGGTTTTGTT GTTGGTCTAA TTGCCAATTC TGTTGGTTCA 6840

GGGTTTATCG GTGGTATCTT GGGAGGTTAT ATAGCTGGTT TCTTGGTTCA AGCGATTATT 6900

AAAAAGGTCA AAGTACCAAA CTGGATTAAA GGTTTAATGC CAACCTTGAT TATTCCTTTT 6960

GTAGCCTCTT TGGTAAGTAG TTTGATTATG ATTTATATTA TTGGAGCGCC TATCGCAGCC 7020

TTTACCAACT GGTTGACGAG CTTATTACAA AGCTTGGGAA GTGCTTCAAA TGGTTTGATG 7080

GGGGCAGTTA TTGGAATTCT CAGTGCTGTT GACTTTGGTG GCCCACTTAA TAAAACAGTC 7140

TATGCGTTTG TGTTGACTTT ACAGGCTGAA GGTGTGAAAG AACCATTGAC TGCTTTACAA 7200

TTGGTGAATA CTGCTACACC AGTTGGATTT GGATTGGCCT ATTTTATCGC GAAATTACTC 7260

AAAAAAAATA TCTATACTCA AGAGGAAATC GAAACATTGA AATCGGCTGT TCCTATGGGG 7320

ATTGTCAATA TTGTTGAAGG TGTAATTCCG ATTGTTATGA ATAACTTGGT TCCAGGTCTC 7380

ATTGCAACAG GTATCGGTGG TGCTGTTGGT GGTGCTGTTT CTTTGACAAT GGGTGCTGAT 7440

TCTGCTGTGC CATTTGGTGG AGTGCTTATG TTACCAACCA TGACTCGTCC AGTAGCTGGT 7500

ATTTGTGCCT TGTTAGCTAA CATTGTAGTC ACAGGACTTG TCTACGCGAT TTTGAAAAAA 7560

CCAATAAAAC ATGCAGAACC AGTTATGACT GTTGAAGAAG AGATTGATTT GTCAGATATT 7620

GAAATTTTGT AAGAGGGTAA CGATGTCAAG AATTGAATTT TCACCATCTT TGATGACCAT 7680

GGATTTGGAC AAATTCAAAG AGCAGATTAC TTTTTTGAAT GATAAAGTAG CATCTTATCA 7740

TATCGATATT ATGGATGGCC ATTTTGTTCC CAATATTACC TTGTCTCCTT GGTTCATTCA 7800

AGAAGTTCAA AAAATTAGTG ACACACCTTT ATCAGTTCAT CTGATGGTCA CAGACCCAAC 7860

CTTTTGGGTA GATCAAGTTC TCGATTTACA ATGTGAGTAT ATTTGTATTC ATGCTGAAGT 7920

TCTGAATGGT CTTGCTTTTC GTTTGATTGA TAAAATTCAT GATGCAGGTC TAAAGGCTGG 7980

TGTTGTCCTT AATCCTGAAA CACCTGTTTC TACAATCTTT CCCTACATTG ATTTACTTGA 8040

CAAAGCAACT ATTATGACTG TAGATCCAGG TTTTGCAGGA CAACGCTTTT TGGAGTCTAC 8100

CTTGTATAAA ATCCAAGAAC TCCGTCAGCT TAGAGTTCAG AATGGTTATC ACTACATCAT 8160

TGAGATGGAT GGTTCTTCGA GTCGTAAGAC TTTCAAACAA ATTGATGTGG CAGGACCAGA 8220

TATTTATGTT ATAGGTCGCA GTGGATTATT TGGTTTGGAT GACGATATTG CCAAAGCCTG 8280

GGATATCTGT TCTAGAGATT ACGAAGAAAT GACCGGAAAA ACAATGCCAA TCAAATAATG 8340

GTTTGAGAAG AAATTTATTA GTTAGGAGGA ATATATGTCA CTACAATCAG TTAACGCCAT 8400

TCGTTTTCTT GGCGTAGATG CTATTAACAA ATCTAATTCT GGTCACCCGG GAATTGTCAT 8460

GGGTGCTGCG CCAATGGCTT ATAGCCTATT TACAAAGCAC CTTAGAATTA CACCTGAGCA 8520

GCCAAACTGG ATTAACCGAG ATCGCTTTAT CTTGTCTGCG GGTCATGGAT CAATGCTACT 8580

GTATGCTCTC TTGCATTTAA CAGGGTATAA GGATGTATCC ATGGACGAGA TTAAAAATTT 8640

CCGGCAATGG GGATCTAAGA CACCTGGTCA TCCTGAAGTG ACGCATACGT CTGGTGTGGA 8700

TGCGACATCT GGTCCGCTTG GTCAGGGGAT TTCTACTGCC GTTGGTTTCG CCCAAGCAGA 8760

GCGTTTTTTA GCTGCTAAGT ACAACAAAGA TGGTTTCCCT ATTTTTGACC ATTATACTTA 8820

TGTTATCGCT GGAGACGGTG ACTTCATGGA AGGAGTGTCT GCGGAGGCGG CTTCTTATGC 8880

AGGTCATCAA GCTTTAGATA AGCTTATCGT CCTCTACGAC TCCAACGACA TCTGCTTGGA 8940

TGGTGAGACC AAAGATACTT TCTCTGAAAA TGTTCGCGTC CGTTACGATG CTTATGGTTG 9000

GCATACAGTT CTGGTAGAAG ATGGAACAGA TTTAGCAGCA ATTTCTACAG CAATTGAGAC 9060

GGCCAAGTTT TCTGGTAAAC CGAGTTTGAT TGAAGTGAAA ACGGTAATTG GTTACGGCTC 9120

ACCCAATAAA AGTGGTACAA ATGCTGTTCA TGGTGCACCA CTAGGAGCAG AAGAAACAGG 9180

AGCAACTCGT AAGTTTTTGG GATGGGATTA CGATCCATTT GAAGTACCAG AGGAAGTATA 9240

TTCTGATTTC AAGACAAATG TAGCGGATCG TGGTCAGGAG GCATACGATG CTTGGGCTAG 9300

TTTGGTGTCT GATTACAAGG TTGCTTATCC CGAAGTTGCT AGTGAGATTG ACGCTATTGT 9360

AGCTGGAAAA TCCCCTGTAA CCATTACTGA AAAAGACTTC CCTGTCTATG AGAATGGCTT 9420

CTCTCAAGCA ACTCGTAATT CGTCCCAAGA TGCTATTAAT ACAGCAGCAG TTTTACCAAC 9480

CTTCTTAGGT GGATCGGCAG ACTTAGCTCA CTCTAACATG ACCTACATCA AGGCAGATGG 9540

CTTACAAGAT AAATATAATC CATTAAACCG CAATATTCAG TTTGGGGTAC GTGAATTTGC 9600

CATGGGAACA ATCCTCAATG GAATGGCTCT TCATGGTGGT TTACGAGTTT ATGGCGGAAC 9660

CTTCTTTGTT TTCTCTGACT ACGTCAAAGC TGCTATTCGG CTATCAGCCA TTCAGGAGTT 9720

GCCTGTAACT TATGTCTTTA CCCATGATTC AATTGCCGTT GGTGAAGATG GTCCAACTCA 9780

TGAACCAGTT GAACATTTGG CAGGTTTACG CTCAATGCCA AACTTGACTG TTATCCGTCC 9840

AGCGGATGCC CGTGAAACTC AAGCGGCTTG GCATCATGCC TTGACCAGTA CCACCACTCC 9900

AACTGTCATT GTCTTAACCC GTCAAAACTT GGTAGTTGAA GAAGGGACAG ACTTTGGTAA 9960

GGTCGCTAAA GGAGCCTACG TCGTGTATGA TACCCCGGGA TTTGATACTA TTATCATTGC 10020

TACAGGATCT GAGGTCAATC TAGCTATCAA AGCTGCTAAG GAATTGGTTT TACAAGGTGG 10080

TAAAGTACGT GTGGTATCTA TGCCCTCAAC CGAACTATTT GATGCTCAAG ATGCTACCTA 10140

CAAGGAAGAC ATTTTACCAT CTAAGACTCG TCGTCGTGTG GCCATTGAAA TGGCAGCGAC 10200

CCAAAGTTGG TACAAGTATG TTGGTTTGGA TGGCGCGGTC ATCGGTATTG ACATCTTCGG 10260

TGCGTCTGCC CCAGCTCAGA CTGTGATTGA TAATTATGGA TTTACGGTAG AGAATATCGT 10320

TGCTCAAGTT AAGTCCCTAT AGAAACCAAT TACAATGAAG ATACAGCTGT TGTCAGACTA 10380

GCAGATGTAG TGATAGACAC TAATCAGATG ATTGGTTATT TAAAAACTGT AATGAAAATG 10440

TAATAATTTA TCTACGAAAG TTATAGTAGA TAGTATACAC AATAGAGTAT ACCCTGAAAC 10500

GGTTGCGAAG TACGCTAATC ACTTTGCTAC TGATCTAGAT AGTTTCTTTA ATCAATAAAC 10560

ACAGCATCCA CAGATTGACT TAGGATATTG TAAGTTTTTT GAAAGCTAGA GAGAAGGTCT 10620

CTAAAATTAA AAAACGCATA GTATAGGATG TTGAAATGAT GAACTGCACC CCAAAAGTTA 10680

GACAGAAAAA AATCTAACTT TTGGGGTGTT TTTATTATGA AATTAACTTA TGATGATAAA 10740

GTTCAGTTCT ATGAACTTAG AAAACAAGGA TATATCTTAG AGAAGCTTTC AAATAAATTT 10800

GGGATAAATA ATTCTAATCT TAGGTACATG ATTAAATTGA TTGATCGTTA CGGAATAGAG 10860

TTCGTCAAAA AAGGGAAAAA TCGTTACTAT TCTCCTGATT TAAAACAAGA AATGATTCAT 10920

AAAGTCTGAC ATGAAGGCTG GACTAAAGAT AGAGTTTCTC TTGAATACGG TCTCCCAAGT 10980

CGTACGATAC TTCTTAACTG GCTAGCACAA TACAGGAAAA ACGGGTATAC TATTGTTGAG 11040

AAAACAAAAG GGAGAGTACC TGAGAGCGGA GAATGCCATC CTAAAAAAGT TAAGAGAACT 11100

CCGATTGAAG GAGGAAAAAG AGAAATAAGA AAGACAGAAA TTGTTCAAGA ATTAATGACT 11160

GAGTTTTCGT TAGATCTTCT TCTAAAAGCC ATTAAACTAG CTCGTTGGAC CTACTACTAT 11220

CACTTGAAAC AGCTAGATAA ACCAGATAAG GACCAAGAGC TTAAAGCTGA AATTCAATCC 11280

ATCTTTATCG AACACAAGGG AGATTATGCT TATCGCCGGG TTCATTTAGA ACTAAGAAAT 11340

CGTGCTTATC TGGTAAATCA TAAAAGAGTT CAAGGCTTGA TGAAAGTACT CAATTTACAA 11400

GCTAGAATGC GACAGnAACG AAAATATTCT TCTCATAAAG GAG 11443

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5338 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

CCAATTACAT TATATTATCA AAATCGTCGA AACTGGCTCC ATGAATGAGG CAGCCAAGCA 60

ACTCTTTATC ACTCAGCCAA GTCTCTCCAA TGCAGTGCGA GATTTGGAAA ATGAAATGGG 120

CATTGAGATC TTTATCCGCA ATCCCAAGGG AATCACCTTG ACCCGTGATG GCATGGAGTT 180

TCTCTCTTAT GCCCGTCAGG TTGTCGAGCA GACCCAGCTT CTGGAGGAAC GCTATAAAAA 240

TCCTGTCGCC CACCGCGAAC TCTTTAGCGT TTCGTCTCAA CACTATGCCT TTGTGGTCAA 300

TGCCTTTGTC TCTTTGCTCA AGAAAAGCGA TATGGAGAAA TACGAACTCT TCCTTCGTGA 360

AACTCGGACT TGGGAGATTA TCGACGACGT CAAGAACTTC CGCAGTGAGG TCGGGGTCCT 420

CTTCTTAAAC AGTTACAACC GTGATGTTTT AACCAAGATG CTGGATGACA ATCACCTGCT 480

AGCCCACCAT CTCTTCACAG CGCAACCGCA TATCTTTGTC AGCAAGACCA ACCCTCTGGC 540

AAAGAAAGAC AAGGTGAAAC TGTCTGATTT GGAGAATTTC CCTTACCTCA GCTATGACCA 600

AGGGACGCAC AACTCCTTCT ACTTTTCAGA AGAGATTCTT TCTCAAGAAC ACCACAAGAA 660

ATCCATTGTG GTCAGTGACC GTGCCACCCT CTTTAATCTC TTGATTGGTT TGGATGGTTA 720

TACCATTGCG ACAGGGATTT TGAACAGCAA CCTAAACGGA GACAATATCG TTTCTATCCC 780

ACTGGATATT GATGACCCGA TCGAGCTGGT CTATATCCAG CATGAGAAAA CCAGCCTATC 840

TAAGATGGGC GAACGCTTTA TAGACTATCT CCTAGAAGAA GTTCAGTTTG ATAGTTGAGA 900

AATGATAAGA ACCAATATGT AGGCTAGCAA CAACCTGCAC ATTGGTTCTT TTTACTTATA 960

ATTAAAAGTT TCCCCTGCCA ACTTATCAGC TAGCTTGGGA AAGAGAGTAT AAAACTTATG 1020

GGCTAGGTTC AACAAAATCG GGAGATTGAG TTCTCGTTTG TTTTTTCCTA TAATCTTGAC 1080

AATCTTTTTA GCCACTGCAT CTGGTTCTAG CAGGAAGCGA TCAACCGATT TAAGATAAGT 1140

TCCATCTGGG TCGGCTTGGT CGAAAAATCC TGTACGGATT GGTCCTGGAT TGACTGTTGT 1200

CACATAGACT CCATAGGGCA TAAGTTCGAG TCGCAGAGCA TTTGAAAAAC CAATAGCCGC 1260

AAACTTGGTC GCTGAGTAAA GACTAGACTT GCCAGTAGCT ATTAGACCTG CCATGCTGAC 1320

GATGTTGATG ATATGCCCTT TGCTGCTTTC CTTCATACGA GCCGCAAGGT GACGAGACAG 1380

ATTCATCAGG GCAAAGGTAT TGACCTCAAA CATCTGGTGA ATATCTTTAT CAGCAATCTG 1440

GTCAAATCCC TCAAAAATCC CGTAACCAGC GTTGTTAATC AAGACATCAA TCTTGCCATA 1500

GCGGAGATAA AGATCAGTTA CCAGAGCTTC TAGGGCTGAA TCGTCGGTAA TATCAATTTC 1560

AATCAATTCT GCATGGGAAT AATTTCCGTA GAGTTGGGCT AATTTTTCCT TATTTCTACC 1620

AAGCAAGATG AGTTGGTCAT TGGGCAGGAG TTTGACCATT TCTTGAGCTA GACCACCGCT 1680

AGCTCCGGTA ATGAGAATAG TAGGCATACT TATCCTTTCT GTGACTGCTA GATTTCCACT 1740

TCTTCCAAGT CTTTGACCAC ATGGACATTT TCAAAAATTG TGGCAGCGTC TTTCTTGAGT 1800

TTGCTAATAT CTTTTGAGAG GAAACGGGCA CTGATATGGT TGAGTAGGAG GCGTTTGGCA 1860

CCTGCTTCTA CCGCTACTTG TGCAGCTTGC ATATTAGTTG AGTGACCATG GTTACGAGCA 1920

ATTTTTTCAT CACCCTTGCC ATAAGTGGAC TCATGAACTA GGACATCTGC ATTGACAGCC 1980

AGACGCACAC TGGCACCCGT TTTTCGAGTG TCTCCTAAAA TAGTGATAAT CTTACCTGGA 2040

CGTGGCGCTG AGATATAGTC TGCTGCCTTG ATTTCAGTTC CGTCTTCCAA AACAAGATCC 2100

TGGCCGTTTT TGATTTTACC AAAAAGCGGG CCGAACGGAA CACCAGCAGC CTTGAGTTTT 2160

TCAGCATCCA GCGTCCCTTC TAGATCCTTT TGCATGACAC GATAGCCAAC ACAGAAAATA 2220

GTGTGGTCCA GCTCCTCTGC ATACACAGTG AATTTATCGG TTTCAAGAAT TTTACCCAGA 2280

GAATCTTGGT CAAACTCATG GAAATGAATG CGGTAGGGCA GACGAGAACC TGACACACGA 2340

AGGCTGGTTA AGACAAATGA CTTGATTCCT TGAGGTCCGT AGATTTCCAA ATCTGTCTGC 2400

TCTTCATTGG CCTGAAAGGC ACGGCTAGAA AGGAAACCTG GCAAACCAAA AATGTGGTCT 2460

CCATGCAGAT GGGTAATAAA GATTTTGCTG ACCTTACGTG GTCGAATTGT GGTTTCCAGA 2520

ATGCGATTTT GCGTACCTTC TCCACAGTCA AAGAGCCAAA CTTCGTTAAT CTCATCCAAA 2580

AGTTTCAGGG CGAGACTTGA AACGTTGCGG GCTTTAGAGG GCTGACCAGC CCCCGTTCCT 2640

AAAAATTGAA TATCCATTCG ATACTTTCTA ATTAATCAAT ATATAACATG GCTGTGCGGT 2700

TTTCCGATCG GAAATAGCGT TTGCCAGAAA AAGCAGCAGC TTCTTGCAAT AAATCCTCTT 2760

GGCTGTAGCC TTTGAGACGT TTTCGACCAT CAGCCAATCT TTCCAAATCA GTCAAAGCTG 2820

TGAGACTTTC TAGGCTGATA ACTTCCTCGT CCTCGACAGG CTTCATGTAA ATCTTACCAG 2880

ACTCTTCAAA GACTAATTGA TGGGGGAAAA TTTGCGCAAT TTCAAAGAGC AAGTCATCCG 2940

AGATTTTCTC CTCATTTTCA AAGAAAATCC GACCAAGGCC GTCACTCTCA TAACAAAAAC 3000

CAAAGGATTT ACCAGACAGA TTAAGCCGAA TAAAAGGCTT ATTTTCTAGG GTGAAACTTG 3060

GCTCAGTATT GTAAAGATTC AGTTCCTGAC TGAGTTCTGC AAAATAATCC GTCGCAGCCT 3120

GAGGACTCTT TTTCTGATAG AGTTCTGCAA AGTAGGCATT AACAACACTT GGCGGAGGTG 3180

TAATAAGTGT TAACTGCTCC TGATCTGTTT TACCAGCTAG AAGCTGATCC AGATAGACCT 3240

TGTCCAGACT TGTATAACCT CCATACTTTA GAGCCAAAGT TTTAATATCA GTCATAAAAT 3300

TCTTCTAACC TCCATTTATT TTTCTCGGAA ATGTAGCCTG TAATC..TTC GCCGTCTTCC 3360

TGATAATCAC GTTCTTCCAG AATTGCAACA CTCTCTAAAT CATGAATCTT GTAGGACTTT 3420

GAAAAAGGCA CTCGCAGGGT AAATGCTTCA AAAATTTCCT TAATCTTATC TAGCAATAAT 3480

GCTTGCAAGT TTTCACGACT GTCCTCAGAC TTGGCAGAAA TGAGGGTATA TGGCGTTTGG 3540

GTAGGCGTGA AATCCTCCAC CAAATCCGCT TTATTATAAA GCGTCAAGTG AGGAATATCT 3600

TCCATGTCCA GGTCTTTCAT GATGGAGAGA ACCGTTTTTT CATGCTCCTC GTGGTAAGGA 3660

TTGCTAGCAT CGATAACATG AACCAGAAGG TCCACATGCT TGCTTTCTTC CAAGGTTGAC 3720

TTGAAACTGG ACACCAACTC TGTCGGCAAA TCTTGGATAA AGCCAACGGT ATCTGTCAAA 3780

GTTACTTGGA GATTGCCTCC CAGATGAATA CTCTTGGTTG TCGCATCCAG AGTCGCAAAG 3840

AGCTCATCTG CTTCATACTG GGTCTTACTG GTCAAGATGT TCATGATAGT TGATTTCCCA 3900

GCATTAGTAT AACCAATCAA ACCAATCTTA AAAGTGCTAG ACTCCAAACG TTTTTCTCTG 3960

ACAGTCGCAC GATTTTTCTC AACCACCTTG AGCTGGCGCT CGATATCCGT GATTTGATTG 4020

CGAACGCTAC GACGGTTCAG CTCCAGCTGG CTTTCACCAG GACCACGGGA ACCAATTCCC 4080

CCTgCCTGAC GGCTGAGCAT AATCCCCTGA CCAACCAAGC GAGGCAAAAG GTATTTGAGT 4140

TGGGCTAGGT GGACTTGGAG CTTCCCTTCA TGGCTTCGAG CCCGCATGGC AAAGATATCC 4200

AAAATCAACT GCATACGGTC AATGACCTTA ACACCGAGAA CTTCCTCTAG ATTGACATTC 4260

TGCCTTGGGG TCAGACGATT GTTGACGATG ACAGTAGTGA TTTCTTCTGC ATCCACCATA 4320

AGCGCAATCT CTTCCAACTT ACCAGAGCCG ACGAAGGTCT TGGAATCATA TTTTTCACGT 4380

TTTTGTCTGT AGCTATCTAC AACGACTGCC CCTGCCGTTT TCGCTAAACT AGCCAATTCT 4440

TCCATGGAGA GGTCAAAACT GTCCATACCC TGCAATTCCA CACCAATCAG CAGGACTCGC 4500

TCCTCTTTTT TCTCCGTTTC AATCATCTAA AAACTCCTCT ATCTGGCTTA AAATGCGGTC 4560

TTGTACACCA GATTCTCCAA TCTGATAAAA GGTGACCTGC ATGCGATTAC GGAACCAGGT 4620

CAGCTGACGC TTGGCAAAAC GACGAGTCGC CTGTTTAAGA CTCTCACTAG CTTCCTCCAA 4680

GGTCTGCTCT CCACGGAAAT AAGGAAAGAG TTCCTTATAG CCAATTCCTT TAGCAGCCTG 4740

TACATTAGGG GAATGGTCAA ACAGCCACTT GGCCTCATCC AAAAGCCCAG CCTCAAACAT 4800

CAAATCCACT CGGTGGTTGA TACGCTCATA AAGTTGACTA CGTTCATCAT CCAAGCAGAT 4860

AATCAGCGGT TCATACAAGG TCTCTTGATT TTCCAAATCC TGACCAAAAT GGGCAATTTC 4920

TAAGGCACGC ATAGCACGAC GACGATTAAA CTGGGGAATC TCAAGGCCTG CTTGATCCAC 4980

CAAATGGGCT AATTCCTCAT CTGAATATGG CTCCAAACTA GCTCGATAAG CTAAAATCTC 5040

CTCATGAGGA GTCTCCCCAC CTAGGTGGTA ACCTTCTAGC AAGCTCTGGA TATAAAGTCC 5100

AGTCCCACCG GCGATAATGG CTAGCTTGCC ACGGTTGTGA ATACCCTCAA TAGTCATCTT 5160

AGCTTCTGAA ACAAAATCAA AAGCCGAGTA AGACTCGGTT ATCTCTCTAA CATCGATTAA 5220

ATGATGAGGA ACAGCTGCCT GCTCTTCTGG ACTAGCCTTG GCCGTCCCAA TATCAAGTCC 5280

TCGATAGACT TGCTGGCTAT CTCCACTAAC CACTTCGCCA TTAAAACGCT TTGCGGGG 5338

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19446 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

CGGAAACCCA TCTAGTCTCC ATCGTTTGGG AGACCAAGCA ACACGAATCT TAGATGCTTC 60

TCGCCAACAG ATTGCAGATT TAATCGGTAA GAAAAGCGAT GAAATCTTCT TTACCTCGGG 120

TGGAACAGAA GGGGATAACT GGCTTATCAA GGGTGTGGCC TTTGAAAAAG CTCAGTTTGG 180

CAAGCACATC ATTGTTTCAG CCATTGAACA TCCAGCAGTC AAAGAGTCAG CCCTCTGGTT 240

GAAAAGTCAA GGATTTGAAG TGGATTTTGC TCCAGTTGAT AAGAAAGGCT TGGTCGATGT 300

TGAGGCGTTA CAGGTTTGAT ACGGCATGAT ACAATCCTCG TTTCCATCAT GGCTGTGAAC 360

AATGAAATCG GCTCTATCCA ACCTATTGAG GCTATTTCAG AATTCTTGGC AGACAAGCCG 420

ACTATTTCCT TCCACGTTGA TGCGGTTCAG GCGCTTGCCA AAATTCCGAC TGAAAAGTAT 480

CTGACAGAAC GGGTGGATTG CGCGACTTTC TCTAGTCACA AGTTCCACGG GGTTCGAGGT 540

GTTGGCTTTG TCTATATCAA ATCTGGCAAG AAGATTACAC CTCTTCTTAC AGGTGGTGGC 600

CAGGAGCGAG ATTATCGTTC GACAACTGAA AATGTGGCAG GGATTGCAGC GACAGCCAAG 660

GCCCTCCGTT TGTCTATGGA AAAGCTAGAT ATCTTTAGGA GCAAGACTGG GCAGATGAAG 720

GCAGTGATTC GCCAAGCTCT TCTGAACTAT CCGGATATTT TTGTCTTTTC AGATGAGGAA 780

AACTTTGCAC CTCATATTCT GACTTTTGGA ATCAAAGGTG TTCGAGGTGA AGTCATCGTT 840

CACGCCTTTG AAGACTATGA TATTTTCATC TCAACAACCT CAGCTTGTTC ATCTAAGGCA 900

GGAAAACCAG CCGGTACCTT GATTGCCATG GGAGTGGACA AAGATAAGGC CAAGTCAGCT 960

GTGCGTCTTA GCCTAGACTT GGAAAATGAT ATGAGTCAGG TCGAGCAGTT TTTGACCAAG 1020

TTAAAATTGA TTTACAATCA AACTAGAAAA GTAAGATAGG AGCATTCATG CAGTATTCAG 1080

AAATTATGAT TCGCTACGGA GAGTTGTCAA CCAAGGGTAA AAACCGTATG CGTTTCATCA 1140

ATAAACTTCG TAATAATATT TCGGACGTTT TGTCTATCTA TACCCAAGTT AAGGTAACAG 1200

CAGATCGCGA CCGTGCCCAC GCTTACCTCA ATGGAGCTGA TTACACAGCA GTTGCAGAAT 1260

CTCTCAAACA AGTTTTTGGA ATTCAAAACT TTTCTCCTGT TTATAAGGTT GAAAAATCTG 1320

TAGAAGTTTT GAAGTCTTCT GTCCAAGAGA TTATGCGGGA CATCTACAAG GAAGGTATGA 1380

CCTTTAAGAT TTCTAGCAAG CGTAGCGACC ACAACTTTGA ACTTGATAGT CGTGAACTCA 1440

ACCAAACACT TGGAGGGGCT GTATTCGAAG CCATTCCAAA TGTGCAAGTT CAAATGAAAA 1500

GTCCTGACAT CAATCTTCAG GTGGAGATTC GTGAAGAAGC AGCCTATCTT TCTTATGAAA 1560

CCATTCGTGG GGCTGGTGGT TTGCCAGTTG GAACTTCAGG TAAAGGGATG CTCATGTTGT 1620

CAGGAGGGAT TGACTCACCT GTAGCAGGTT ATCTTGCTCT TAAGCGTGGG GTGGATATCG 1680

AGGCAGTTCA CTTTGCTAGT CCACCATATA CTAGTCCTGG TGCCCTCAAG AAAGCGCAGG 1740

ACTTGACCCG TAAATTGACC AAGTTTGGCG GAAATATCCA GTTTATAGAG GTGCCTTTCA 1800

CAGAGATTCA AGAGGAAATC AAAGCCAAAG CGCCAGAAGC TTATTTGATG ACTCTAACTC 1860

GTCGCTTTAT GATGCGGATT ACTGACCGTA TTCGTGAGGT ACGAAATGGT TTGGTTATCA 1920

TCAATGGGGA AAGTCTAGGT CAAGTAGCCA GCCAAACCCT TGAAAGTATG AAGGCTATCA 1980

ATGCTGTTAC CAACACTCCC ATCATTCGTC CTGTGGTTAC CATGGACAAG TTGGAAATCA 2040

TTGACATCGC CCAGGAAATC GATACCTTTG ACATTTCAAT CCAACCGTTT GAAGACTGTT 2100

GTACCATTTT TGCACCAGAT CGTCCAAAAA CAAATCCTAA AATTAAGAAT GCGGAGCAGT 2160

ACGAAGCGCG TATGGATGTT GAAGGCTTGG TTGAGCGAGC AGTGGCTGGA ATCATGATTA 2220

CTGAAATCAC ACCTCAAGCC GAAAAAGATG AAGTTGATGA CTTGATTGAC AATCTGCTCT 2280

AATTCAGAAA ATCCAAAAGA ATAGCGAAAA TCAGTAAAAA AAGTTAGTTT TTTCTCTAAA 2340

AACAGGTAAA AAACTAACTT TTTTTATTTT TATGATATAA TGATATAAAA TTTTGAATAT 2400

AGAGAGTTTT CTGACAATGA ATCAATCCTA CTTTTATCTA AAAATGAAAG AACACAAACT 2460

CAAGGTTCCT TATACAGGTA AGGAGCGCCG TGTACGTATT CTTCTTCCTA AAGATTATGA 2520

GAAAGATACA GACCGTTCCT ATCCTGTTGT ATACTTTCAT GACGGGCAAA ATGTTTTTAA 2580

TAGCAAAGAG TCTTTCATTG GACATTCATG GAAGATTATC CCAGCTATCA AACGAAATCC 2640

GGATATCAGT CGCATGATTG TCGTTGCTAT TGACAATGAT GGTATGGGGC GGATGAATGA 2700

GTATGCGGCT TGGAAGTTCC AAGAATCTCC TATCCCAGGG CAGCAGTTTG GTGGTAAGGG 2760

TGTGGAGTAT GCTGAGTTTG TCATGGAGGT GGTCAAGCCT TTTATCGATG AGACCTATCG 2820

TACAAAAGCA GACTGCCAGC ATACGGCTAT GATTGGTTCC TCACTAGGAG GCAATATTAC 2880

CCAGTTTATC GGTTTGGAAT ACCAAGACCA AATTGGTTGC TTGGGCGTTT TTTCATCTGC 2940

AAACTGGCTC CACCAAGAAG CCTTTAACCG CTATTTCGAG TGCCAGAAAC TATCGCCTGA 3000

CCAGCGCATC TTCATCTATG TAGGAACAGA AGAAGCAGAT GATACAGACA AGACCTTGAT 3060

GGATGGCAAT ATCAAACAAG CCTATATCGA CTCGTCGCTT TGCTATTACC ATGATTTGAT 3120

AGCAGGGGGA GTACATCTGG ATAATCTTGT GCTAAAAGTT CAGTCTGGTG CCATCCATAG 3180

TGAAATCCCT TGGTCAGAAA ATCTACCAGA TTGTCTGAGA TTTTTTGCAG AAAAATGGTA 3240

AGTTAAGAAA GGAAAAAACG AAATGCATAT TGAACATCTT AGCCACTGGA GTGGTCATCT 3300

TAACCGTGAA ATGTACCTTA ACCGTTATGG ACATGGTGGG ATTCCAGTTG TGGTCTTTGC 3360

TTCATCAGGT GGTAGTCACA ACGAATACTA TGATTTTGGC ATGATTGATG CCTGTGCTTC 3420

CTTTATCGAG GAAGGCCTTG TCCAGTTCTT TACCCTATCT AGTTTGGATA GTGAGAGCTG 3480

GTTGGCTACT TGGAAAAATG CTCATGACCA AGCGGAAATG CACCGTGCCT ACGAACGTTA 3540

TGTGATTGAG GAGGCCATTC TTTTATCAAG CACAAGACAG GTTGGTTTGA TGGCATGATG 3600

ACGACAGGTT GCTCTATGGG AGCCTATCAT GCACTCAATT TCTTCCTCCA GCATCCAGAT 3660

GTCTTTACCA AAGTGATTGC TCTCAGTGGT GTTTACGACG CACGTTTCTT TGTCGGTGAT 3720

TACTACAACG ATGATGCTAT TTACCAAAAC TCGCCAGTAG ATTATATTTG GAACCAAAAC 3780

GACGGCTGGT TTATTGACCG TTACCGTCAG GCAGAGATTG TGCTGTGTAC GGGGCTTGGA 3840

GCCTGGGAAC AAGATGGTTT GCCATCCTTT TACAAGCTCA AAGAAGCCTT TGACAAGAAA 3900

CAAATTCCAG CCTGGTTTGC TGAATGGGGA CATGATGTCG CCCATGACTG GGAATGGTGG 3960

CGTAAACAAA TGCCTTATTT CCTCGGTAAT CTCTATTTAT AAAAGGAGTT ACCTATGAAT 4020

TACCTTGTTA TTTCTCCCTA CTATCCACAA AACTTTCAAC AGTTTACCAT CGAACTAGCT 4080

AATAAAGGCA TCACAGTCTT GGGAATTGGT CAAGAGTCTT ACGAGCAATT GGATGAGCCC 4140

TTGCGCAATA GCTTGACCGA GTATTTTCGT GTTGATAATC TTGAGAACAT AGATGAAGTC 4200

AAACGTGCAG TTGCTTTTCT CTTTTATAAA CATGGTCCAA TTGGCCGCAT CGAGTCTCAC 4260

AATGAATACT GGCTTGAGCT AGACGCAACA CTCAGAGAAC AATTCAATGT TTTTGGTGCC 4320

AAACCAGAGG ATCTCAAAAA GACGAAATAT AAGTCTGAAA TGAAGAAACT TTTCAAAAAA 4380

GCAGGTGTTC CTGTGGTACC TGGAGCTGTT ATCAAGACGG AAGCAGATGT TGATCAAGCA 4440

GTGAAAGAAA TCGGTCTTCC AATGATTGCC AAACCTGATA ATGGAGTGGG AGCAGCCGCA 4500

ACCTTTAAAC TTGAGACAGA AGACGATATC AATCACTTCA AGCAAGAATG GGACCATTCA 4560

ACCCTTTATT TCTTTGAAAA ATTTGTCACT TCCAGCGAAA TCTGTACCTT TGACGGGCTC 4620

GTGGACAAGG ATGGAAAGAT TGTCTTCTCA ACAACCTTTG ACTACGCCTA TACACCGCTT 4680

GACCTCATGA TTTATAAGAT GGACAATTCT TATTATGTGC TCAAGGATAT GGATCCTAAA 4740

CTGCGCAAGT ATGGGGAAGC AATTGTCAAA GAATTTGGTA TGAAAGAACG GTTTTTCCAT 4800

ATTGAGTTCT TCCGTGAGGG GGACGATTAT ATTACCATCG AGTACAATAA CCGCCCTGCA 4860

GGTGGTTTTA CCATTGATGT TTATAACTTT GCTCATTCCT TGGACCTTTA TCGTGGCTAT 4920

GCAGCTATTG TCGCAGGAGA GGAGTTCCCG GCGTCAGACT TTGAAACTCA GTATTGTTTG 4980

GCTACTTCTC GCCGTGCAAA TGCTCACTAT GTTTATTCAG AAGAGGATTT GCTTGCCAAA 5040

TATAGCCAGC AGTTCAAGGT TAAAAAAGTC ATGCCAGCTG CCTTCGCGGA ACTTCAAGGA 5100

GATTACCTGT ATATGCTGAC CACTCCGAGT CGACAAGAAA TGGAGCAGAT GATTGCAGAT 5160

TTCGGACAAC GTCAAGAATA AGAACTATCG GATTAAGGAA ATTAACTCCC TTAATCCTTT 5220

TGTTTTGTCT GATAAAAAAT AAGAGCATCC CAACAAGGTA GCTATCATAA AACTTGTTCG 5280

ATAACTATTT GAAGCAGGAT TAGGTGGTCA GAAATTAAAT TTTAATATTT CAATTGAGTC 5340

ATAGTATTGT GTTTGCGTAT CCTTAAATCA GCTAAAAGGA TCCATGACGA CACCTATACG 5400

ATATAGTTTT CAAGATACCA AACAAGTCTA TTAATATTCA ATGAAAATCA AAGAGCAAAC 5460

TAGGAAGCTA GCCGCAGGTT TCTCAAAACA CTGTTTTGAG GTTGTGGATA GAACTGACAG 5520

AGTCAGTATC ATATACTACG GCAAGGTGAA GCTGACGTGG TTTGAAGAGA TTTTCGAAGA 5580

GTATAAAATA TTCAGGTGAC GCATAGATAT AGTTAATTGA AGCTTTGTTT GAAATCTGAT 5640

AAAATAATGA TATTACTAAG TTTTAAAAAC TAAAGAAAAG GGAAGATATG ATTACAGGCG 5700

AATTAAAAAA TAAAATCGAT CAGCTGTGGG AAATTCTTTG GACAGAAGGA AACGCAAATC 5760

CTTTAACAAA TATTGAACAG TTGACTTATC TCTTATTTAT GAAAGATTTG GATAGTGTCG 5820

AGCTTGGACG TGAAAGTGAT GCTGAATTTC TAGGGATTCC TTATGAGGGA GTTTTTCCAA 5880

AAGATAAACC TGAATACCGT TGGTCAACTT TTAAAAATAT AGGAGATGCT CAGGAAGTTT 5940

ATCGTTTAAT GACTCAGGAG ATTTTTCCGT TTATTAAAAA TCTCAAGGGG GATACAGATG 6000

ATACAGCCTT TTCACGATAT ATGCGAGAAG CTATTTTTCA AATAAATAAA CCTGCTACGC 6060

TTCAAAAGGC AATTTCTATC TTAGATGTTT TTCCAACTAG GGGATTAGAT GTAGATTTTG 6120

ATAATGACAA ACAAAGTATT ACTGATATCG GAGATATCTA TGAATATCTG TTATCAAAAT 6180

TGTCGACCGC AGGTAAAAAT GGACAGTTCC GTACACCTCG TCACATCATC GATATGATGG 6240

TTGAGTTGAT GCAACCGACT ATCAAAGATA TCATCTCAGA TCCCGCTATG GGTTCTGCTG 6300

GCTTCTTAGT ATCTGCTAGC CGTTACTTAA AGCGTAAGAA AGATGAATGG GAAACCAATA 6360

CAGATAATAT CAATCATTTT CATAATCAGA TGTTTCATGG AAATGATACG GATACGACTA 6420

TGTTGAGACT TGGGGCGATG AACATGATGC TACATGGAGT AGAAAATCCA CAAATCAGTT 6480

ACCTTGACTC GCTGTCTCAA GATAATGAAG AAGCCGATAA ATATACTTTG GTTTTAGCAA 6540

ATCCTCCTTT TAAGGGCTCA CTTGACTACA ATTCAACCTC TAATGACCTT CTTGCAACCG 6600

TAAAAACCAA AAAAACAGAA TTACTCTTTC TTTCTCTTTT CTTGCGAACT TTAAAACCAG 6660

GTGGACGAGC AGCAGTTATC GTACCTGATG GTGTCCTTTT TGGTTCGTCT AAAGCTCATA 6720

AAGGAATTCG TCAGGAAATT GTAGAGAATC ATAAGCTTGA TGCTGTAATC TCAATGCCTA 6780

GTGGTGTGTT CAAGCCTTAT GCTGGAGTTT CAACTGCCAT TCTCATCTTT ACAAAAACTG 6840

GTAATGGTGG TACTGACAAA GTCTGGTTTT ACGATATGAA AGCGGATGGT TTAAGTTTGG 6900

ATGATAAGCG ACAACCGATT AGCGACAATG ATATTCCAGA TATTATCGAA CGCTTTCATC 6960

ATCTTGAAAA AGAAGCAGAA CGTCAGAGAA CGGATCAATC TTTCTTTGTT CCAGTTGCTG 7020

AGATAAAGGA AAATGATTAT GATTTGTCTA TCAATAAATA TAAAGAGATT GAGTATGAAA 7080

AAGTTGAGTA TGAACCAACA GAAGTCATAT TAAAGAAAAT CAATGATTTA GAAAAAGAAA 7140

TTCAAGCTGG CTTGGCTGAA TTGGAAAAAT TACTCAAGTA GGGAGGTGGC TGTATGAAAA 7200

AAGTGAAGTT GGGGGAAGTC TTATCTCTAA AAAAAGGCAA GAAAGCCACT GTACTTGCTG 7260

AACAAACAAC TCTAAGCCAA CGTTATATTC AAATAGATGA TTTAAGAAAT AATAATAATT 7320

TAAAATTCAC TGAAAGTTTA AATATGACTG AAGCACTCCC AGATGATATT CTGATAGCAT 7380

GGGATGGAGC TAATGCAGGA ACAGTTGGTT ATGGATTATC GGGAGCTGTT GGTAGTACAA 7440

TTACGGTCTT AAAAAAGAAT GAGCGATACA AAGAAAAAAT TATATCAGAT TACTTGGGAG 7500

TCTTTTTGGA AAGTAAATCG CAGTATTTAC GAGATCATTC AACAGGTGCA ACAATTCCTC 7560

ATTTAAACAA GAATATATTA CTTGATTTAC AATTAGAATT GCTAGGTATC GAAGAACAAG 7620

AGAACATTAT CTGTATTCTT AATACGATTA AAAGGCTTAT TACTAAAAGA AAATTTCAGT 7680

TAGATGAACT AAACTTGCTC GTCAAATCCC GATTTAACGA GATGTTTGGG GAAAATAAAA 7740

TATTTGAAAG CATTGATAAC TTATTTGATA TTATAGATGG TGATAGGGGC AAAAATTATC 7800

CTAAATCAGA TGAGTTGTTT AGTGAGGAGT ACTGTTTATT TTTAAATACA AAGAATGTTA 7860

CTAAAAACGG ATTTTCATTC GATACAAAGC AATTTATCAC TAAAACAAAG GATAAATTAC 7920

TTCGAAAAGG CAAACTTGAG CGTTATGATA TAGTCTTGAC AACAAGAGGT ACTGTTGGAA 7980

ATGTAGCGTA CTACGATGAA TTAATAAAAT ATAAACATTT ACGTATAAAT TCAGGTATGG 8040

TAATATTACG TCCCAAGACA CCAAATCTAA ATCAGAAATT TATTATCCAT GTTTTAAGGA 8100

ATAATAATTA TAGTCGAGTG ATATCAGGAA GTGCTCAGCC TCAGTTACCA ATTACAAAAT 8160

TAAAAAAAAT ACTTCTCCCC CTCCCCCCAC TAGCCCTCCA AAATGAGTTC GCAGACTTTG 8220

TAGTCCAGGT CGACAAATCA CAATTTGCTT GTGAGATAGC TATAAAAGTG TGGAGAAATA 8280

GCTTGAAATT TAGTATAATA TAGCTAAACT ATTTGTTTAA AGTGAGAAAA AAATGGGAAA 8340

TTTTAGCTTT CTTTTAAAAA ATGACGAATA TGAATCTTTT TCAAAACCTT GCATTGAAGC 8400

TGAGAATATG ATTGCTACAT CAACTGTGGC TACTGCCTTT ATGGCGCGTC GTGCTTTAGA 8460

GCAGGCTGTC CATTGGATAT ATAGTCACGA TTCATATTTA GAAGCTCCCT ATCGTGCTAC 8520

TCTATCTTCT TTAGTATGGG ATGATGATTT TAGGGATATC GTAGATTCTG AACTCCACAA 8580

GCAGATAGTT CTGTTGATTC GGTGGGGAAA CCATGCTGCT CATGGTGGTG AAATTAAGGA 8640

ACGAGAAGCG ATTTTAGCTT TGCATCATTT GTATCAGTTT GTTAATTTTA TCGATTATTG 8700

TTACAGCAAT GAGTTTGTGG AGCGTTATTT TGATGAGAAG TGCTTACCAC TTTCAGCAAA 8760

CATCAAATAC CGAGAAACTC CACAATCTAT GATAAAGTTA CAAGACAGTT TACCAGAACT 8820

GCCTGATTTT CATGAACAGA TGGCTGCTCA GTCCGTAGAA GTTCAAGAGA CTTATACTGA 8880

AAAACGTGAG ACTGCAGCGC AACGGCAAGA TGTGCCTTTC CATATTGATC AATTATCTGA 8940

GGCAGAGACA AGAAAGCTCT TTATTGATAT CGATCTCCGT TTAGCAGGAT GGATATTTGA 9000

AGAAAACTGT CGTGTTGAGA TAGCCGTTGA TGGTCTCAAG CACGGTTCAG GAATTGGTTA 9060

CTGTGACTAT GTACTTTATG GTAAAAATGG GAAAATTTTA GCGATTGTGG AGGCTAAAAA 9120

AGCCTCTGTC AATCCAGAAG TAGGGGAAGT ACAGGTCAAA GAATATGCTG AAGCTTTGGA 9180

GAAACATATC GGCTATCAGC CAATTTGCTT TATTACAAAT GGGTTGAAGC ACTATATACT 9240

TGATGGTCCG AACCGCCGCC AGATTGCAGG CTTTTACTCT CAAGAAGAAT TGCAATTAGT 9300

GATGGATAGA CGTCATCTTC AAAAACCGCT TGAGGATATT TCTAGTAAAA TTAGGGACGA 9360

TATTTCCGGG CGTCACTACC AAAAACATGC CATTGCAAGC GTTTGTGAAG CTTTCTCTGA 9420

TCATCGTAGA CAGGCACTTT TGGTTATGGC AACTGGGGCG GGGAAAACTC GTACAGCAGT 9480

TTCTCTAGTT GATATCTTAT CACGTCATAA CTGGGTAAAA AACGTTCTCT TCTTAGCCGA 9540

TAGAACTTCC TTGGTTAAGC AAGCATATGA TTCGTTTAGA AAATTACTCC CAGATCTTTC 9600

CGTTTGTAAC TTCTTAGAAG ATAAAGAAGG AGCTCAATCA AGTCGCATGG TCTTTTCAAC 9660

TTATCCGACC ATGATTGGAG CGATTAGTGG TCAAGAAGAA GTAAATCAAC GCCCTTTCAC 9720

TGTTGGGCAT TTTGACCTTA TCATAATTGA CGAATCTCAC CGTTCTATTT ATCAGAAATA 9780

CAAGTCCATT TTTGATTATT TTGATGCAAG AATTGTAGGC TTAACAGCTA CTCCGCGTCA 9840

AGATTTAGAT AAAAACACCT ATGGATTCTT TAATTTGGAG AATGGGGTTC CAACATATGC 9900

ATATGATTTG GAAGAGGCTG TTAAAGACGG ATATTTAGTA GCCTATCATT CTATCGAAAC 9960

CAAACTGAAA CTACCTACGG ATGGTCTACA TTATGATGAT TTGTCCGAAG AAGAAAAGGA 10020

ACATTTTGAT AGCAAATTTG AAGACAATAG CTGTGAAAAA GATATTGATG GGAGTGTATT 10080

TAATTCCTTT ATTTTCAATA AAAGTACAGT AGAAATTGTT TTAAATGAAC TCATGACAAG 10140

AGGAATTCAG ACAGCCTCGG GTGATGAAAT TGGTAAAACT ATTATTTTTG CTAAAAATCA 10200

TGATCATGCG GAATATATCA GAGGTATTTT TAACAACCGC TATCCTGAAA AAGGGAGCGA 10260

CTATGCTCAG GTGATTGATT ATAGTATTAA GCATTATCAG ACCTTGATTG ATGATTTTAA 10320

AATTAAGGAG AAGTATCCTC AAATTGCGAT TTCTGTCGAT ATGTTAGATA CAGGTATTGA 10380

TGTACCAGAG GTTGTTAATT TAGTCTTCTT CAAGAAAGTA CGCTCTAAAA CTAAGTTTTG 10440

GCAGATGATT GGTCGAGGAA CCCGTCTATG TAAAGATTTA TTTGGACCTG AGCAGGATAA 10500

GGAAAACTTC TTGGTATTTG ATTATGGGGA CAATTTTGAT TATTTTCGTG CAGATCCAAG 10560

AGATGGAGAG GGTCGTCACA TTGTTTCGCT GACTCAGCGT TTATTTAATA TCAAAGTGGA 10620

CTTGATTCGA GAACTTCAGG GACTCCAATA CCAAGAAGAT CAGTTTGCGA GAGCATACCG 10680

TCAGCAGCTT GTCTCGGAAC TTCAAGGTCG TATAGAGAGC TTAAATGAGT TGGACTTCAG 10740

GGTTCGTATG GTTTTAGATA CAGTTTATAG CTATAGGAAA TTGGAAAGTT GGCAGAATCT 10800

AACTGCTGTT ACAAGTGAAA CCATTCAAAA AAATCTCTCT CCGCTTTTAT TTGATGAAGA 10860

TAAAGAAGAT GAGATGGCGA GGAGATTTGA TTTGTGGTTG CTTCATATTC AGTTGGGGCA 10920

ACTGACAGCT AAATCTTCCA CTGTTCATAT TTCCCAAGTG ATGAAGACGG CTAGAGCTCT 10980

TTCTGCTATT GGCAATATCC CGCAGGTTTT TGAGCAGGCT GAAATTATCA GGAAAGTACA 11040

GGAGCCTGAA TTTTGGAAAG AAGTTAACTT GTCTGATTTG GAAAAAATTC GTCTTGCTAT 11100

TCGAGATTTA TTACAGTTTT TGGATAAAAC AGACCGTAAA CCCTACTATG TTAACTTTGA 11160

AGATCGTATA CTCTCCACTG TTCACGAGAC CACAGCATTT TTGCAGGTCA ACGATCTTCG 11220

GTCTTACAAT GAAAAAGTTG AGCATTATTT GAAAACTCAT CTGGATGAGG AGTCCATTTC 11280

TAAGCTATAC CATAATAAAA AGTTGACATC TGATGATATG CTTGCACTTG AAAAATTGCT 11340

TTGGGAAAAA TTAGGTAGTA AAGCAGACTA CCAAAGTCAT TATGAAAATA AGGCAATTCC 11400

GAGATTGGTT CGTGAGATTA TTGGCTTAGA TAGAGAGTCT GCCAATCGTA TTTTTTCTAA 11460

ATTTTTGTCG GATGAGAATC TTAATGCCAG GCAGATTTCA TTTGTAAAAT TGATTGTAGA 11520

CTACATTGTA GAAAATGGTT TTTTAGAGAC GAAAGTGTTA ACGCAAGAGC CGTTTAAATC 11580

TTATGGTTCT GTTCAACTAC TCTTCCAACA CCAACTACCA GTACTTCGTA ATATTGTTCA 11640

AATCATTGAA CTTATCAATA ATCGAGCTGG AGAAGCGGCT TAAATTCTAA AGTGATTGCC 11700

ATGCTGAGAC TCATTTAAAA TTAAAAAGAG TAGAAATTTA TGCTATATAT GAGAAGTTTT 11760

ATTAGGAAGA ATGTCATCGT TTTCCTAGAA TACAGTATCA GTTGTTAAGT GGTTGATAAA 11820

TTTCAAAGTA GATACTTGTA CCACGATGTT TGTTGATCGA GTTATTAACA AAAGAGCTAC 11880

TTTGATTTTA AAGAAATAGA AAACAAAAAG CCGAGCAAGA ATTCAATTGC AGGAGAAAAT 11940

GAAATAATAC TCAATGAAAA TCAAAGAGCA AACTAGGAAA CTAGCTGCAG GCTGCTCAAA 12000

ACACTGTTTT GAGGTTGCAG ATGGAAGCTG ACGCGGATTG AAGAGATTTT CGAAGAGTAT 12060

AAATCTTCCT AGGATAAAGC AAAACGCATA GTATCAAGGG TTTTCAACAC TTGATACTAT 12120

GCGTTTTCTG ATGTTAAAGA CTTTCTACCA GGTTTTTTAA AAGCATAATT GTTAGTTGTA 12180

GTCATTTATT ATTCTTCAAA GAAAAATGGT GGGGCGAATT TTTTCAGTTC TTCAAAGCAC 12240

TTTTGAGCAG TATCTGCATC TTCACAGATG ATAAGACAGA CATCATTACC ACAAAGGGTA 12300

GCGATAGCGT CAGGGAAGCT CAAAGTATCA ATGATAGAAC CAAAGGATTG AGCCAGTCCA 12360

GGAAGGGTTT TTAGTAGGAC TTGGTGTTGA ACTGGGCGCA TCCAGACAAG GGCGTCTTCC 12420

ATGTAGAGTT CGAGACGTTT TTCCCATTTT GAGATGGAAC CATTGTTAAG AACATAATAA 12480

GCGCTATCTT CTTCGCGGAC TTTTGATAGG TTCATATTTT TGATGTCGCG TGAGAGGGTT 12540

GCCTGGGTTA CTTGAATGTC GTTCTCAGCA AGAAGGGCTT GCAACTCAGC CTGTGTATGA 12600

ATCTTGTTTT TTGTGATAAG AGCGCGTATA AGTTGGTGGC GGTGTTCTGA TTTATTCATA 12660

ATAATGTAAC TCCTTTTAGC AAGGTAAGGT AAGCATGGAC TGAGCGAGGT CGACAGTCAA 12720

GTGGTAGTCT GTATTGTCAC GGATGGTGAT TTCAAAGTCA GTAGTATAGA GGACTAAACG 12780

GAGAGTGTCT CCTTCTTTTA GCTTGTAAAT AGTTGGCTGC AGTTCAAATT GAACGTCCAT 12840

CCATTCATCT GCAGTAATAT CCTCTACTAA CAGTAAATCA TTTCTATTTT GTAAATTAAG 12900

GTAACCTTTT GTCACGACTC GTTGTGCCTC TGGTCTAAAT GGCAATTCAC AGAGATTTTC 12960

CAACATGTGA TAGCGACCGT TGTCAATGGT TCTAGCACTT AAAATAGCTG GATAAGGTTG 13020

TAGGTATTTC TTTTGCCCAA ATTCTAGCAG TTGGGCAGAT AAGAGCCCCT TGTTTGTACT 13080

GGATTTGATA CGAAGATTGA GCTGAGCGCG ACCGTTTAGG TGGAGATCTT TAGTCACAGG 13140

AAGGTTAATA GTAATCTGAT TGGCTTTCCC TTGATAGAGC TCTGTATTGA AGGTTTGGTA 13200

TGTCTTACCA TAGCGCTCAA AATCCTTATC TGGGTACTGG TTTTGAATAG CTTGCTCTTC 13260

TTGACCAAGT GAGAAGGTTT CACAGTTTTC TTGCCCACCG AAGTTATCAA GTGATAACCA 13320

AGTCTGTGGA GCTGTATTGT CCTGCCAGAT AACAGTAGGA AGTTGAAAGT CTGTTTCCTG 13380

TCCTAGTAAT TTCTTGGTCA ATAAGGCATT TATGGACTCA GGGAAGTCAA TTGATTGCCA 13440

ATTGTTCATG TAAACATGGG CACCATTATG GAAAAAGAGA TGCTTGTGTA TATGAGTAGG 13500

AAGAGCATGG AACATCTGGT AAACATGAAG TGGTTTGACA TTCCAATCCT GAGAACCATG 13560

AGTAAAGACA ACCTCTGCCT TTACTTTATG GCCATTGAGC AGATAGTTGC GGTCATGCCA 13620

AAACTGATTG TAGTCCCCAG TTTTTCGGTC TAGCTGAGCT TTCACTTTTT CTAAGTCAGC 13680

TTGGTGAGCT TCATTGCCAC GGATATAGTC GCCAGCTAAG AGATTACGAG AATAGGTTAA 13740

CTCAGCAAGG GAGTCAAAGT CCTCACCTGG ATAACCACCT GGGCTAGTCA CCAGACCGTT 13800

TTCACGGTAG TAGTTGTACC ATGATGAAAT TCCTGCCTCG GCAATGATAA CTTCTAAACC 13860

ATCGACTCCT GTAGTCGCAA GACCATTGGA CATGGTACCT AGATAGGAAA GTCCTGTTGT 13920

AGCAACTTTT CCGTTTGACC AATCAGCCTT GACTTGACGC TGGCGCGTGT GATCAGTAAA 13980

GGCACGGCAA CGACCGTTAA GCCAATCGAT GACATTTTTA TAAGCCTCGA TTTGCTGGTA 14040

GTCTCCATTA GTCATGAAAC CTGTCGAGTC TTTGGTACCA ACACCTGAGA CATAGAGATT 14100

GGCAAAGCCT CTCGGAAGGA AGTAGTCGTT TAGTGTATAG CTAGAGTTGA TGTGAGTTAG 14160

CTTTTCCTCA GCCTCTGCTA TAAGCTCAGC TTTACCTTGG GGTTGGACGA GATTTAGTTG 14220

AGGTTTCTCT AGCTCAATCT TGTGAGGAAG CTTAACCTCA AGCTCGCCCT CCATCTTGTA 14280

GAGAGCCTTG TCACTAGCCT TGTCATTGGT TCCCTGATGA TAAGGGCTGG CTGTCATGAT 14340

GGCAGGGATT TTTCCATCAA AACGAGGGCG AATAATGCTA ACCTTTACTA GGTCTGATAG 14400

CCCTTTTTGG TCAGTATCGA CACGAGACTC AACGTAAACG ACTTCACGAA TGACATCCTG 14460

GTTAGAAAAA GTAGCCAAAC TCTTGCCGTT AAAGTAGTGG TAGTCATTAT CCTCCGGAAT 14520

AAGACCATCA CTAACAAGTT GGTCGATAAG AGTATTTCCT TTTTTGGTGC GAGTATTGAG 14580

TAACTGATAG AGATTTTCAA TCAAGTCACC ATATATAATG GGAAATCCAG TTTCTTTACG 14640

AAAAACGTCA CTATCTTCGA AGTCAACCAA ATAAGAAAAG CCTAAAAGTT GAAAAGCAAC 14700

AGTATAAAAA ATATCTGCTG TCAGTTCATC TTCTGATTGA AAAAATGTCA GCAGGTCTGT 14760

TTTTTTATCA GCTGCTAGGA TAGAAAGTGG GTAGTTGGTG TCTTGATAAG TGAAAAAGAA 14820

ACGACGTAAA AAGGTTTCAA GTGAGTCTTT GTGATTGGCT GTATTTTGTA AATCAAAGCC 14880

ACATTTTTTT AGTTCAGATA AGACATTTTC TTTTGGAAAA TTGATATAAC TATATTGATT 14940

AAAACGCATA GAACCTCCAT ATAGAATGAC AGTTAAGGTT ATTATATCAA AAAAAAAGCA 15000

GAAAGGGAAT TGTTAACTTC AAAAGGAAAT AATCCAATAA AAATGAATAA AGTACTAAAT 15060

TCAATATAGA GAACAGAGTA ACAATAAGAA TAAATAGATA GGGTATAAAA GTTCTAGGAG 15120

ATTTATATTA TATGCTTTCT ATTTTTATAT ACAATATAGT ATAAATATAA AAATGATGAC 15180

AAAAATACAA ATGAATAGAA AATAAATTAG TAAGCTGATG AAATTTTTCT CAAGAGAAGC 15240

CATTTATAGG TGAAAATGGT ATAATATAGT GAGAAGGATA GAGGAGAAGT GTAAATTGAT 15300

CGCACAACTA GATACAAAAA CAGTCTATAG TTTTATGGAA AGCGTCATTT CGATCGAAAA 15360

GTATGTGAGA GCAGCTAAAG AATACGGCTA CACTCATTTG GCTATGATGG ATATTGACAA 15420

TCTTTATGGC GCTTTCGACT TTCTAGAGAT TACAAAAAAA TACGGCATTC ATCCTTTGCT 15480

AGGGCTTGAA ATGACAGTGT TTGTAGATGA TCAGGGAGTG AATTTGCGCT TTTTAGCTCT 15540

ATCTAGTGTG GGCTATCAGC AGTTGATGAA GCTTTCGACA GCCAAGATGC AGGGGGAGAA 15600

AACTTGGTCA GTCCTGTCCC AGTACCTGGA GGATATCGCG GTCATTGTGC CTTATTTTGA 15660

TAGAGTTGAG TCGTTAGAAC TAGGCTGTGA TTACTATATA GGGGTTTATC CAGAAACACT 15720

AGCAAGCGAA TTTCATCATC CTATCTTACC TCTTTATCGG GTCAACGCTT TTGAAAGCAG 15780

GGATAGAGAA GTTCTTCAAG TTTTAACAGC GATTAAAGAA AATCTACCGC TCAGAGAAGT 15840

TCCCTTGCGT TCGAGACAAG ATGTCTTTAT ATCAGCAAGT TCTTTAGAGA AACTATTCCA 15900

AGAGCGTTTT CCGCAAGCTT TGGACAATTT AGAAAAGCTT ATTTCAGGCA TTTCTTACGA 15960

CTTGGATACT AGTCTGAAAC TGCCTCGTTT TAATCCAGCT AGACCAGCAG TAGAGGAGTT 16020

GAGAGAGCGT GCTGAACTGG GGCTTGTTCA GAAGGGGTTG ACTAGTAAAG AATATCAAGA 16080

TAGACTAGAC CAAGAATTGT CTGTTATTCA TGATATGGGC TTTGATGATT ATTTCTTGGT 16140

TGTTTGGGAT TTGTTGCGTT TTGGACAATC GAATGGCTAT TATATGGGAA TGGGAAGGGG 16200

TTCTGCAGTA GGCAGTTTGG TTTCTTATGC CTTAGACATC ACGGGGATTG ACCCAGTAGA 16260

GAAAAATCTG ATTTTTGAAC GCTTTCTTAA TCGTGAACGC TATACCATGC CTGATATTGA 16320

TATTGATATC CCAGATATTT ATCGTCCAGA TTTTATCAGA TATGTTGGTA ATAAATATGG 16380

TAGTAAACAT GCGGCACAAA TCGTTACTTT TTCAACCTTT GGAGCCAAGC AAGCTCTTCG 16440

AGATGTCTTG AAACGCTTTG GTGTGCCAGA GTATGAATTA TCTGCAATTA CTAAGAAAAT 16500

CAGTTTTCGT GACAATCTTA AGTCGGCCTA TGAGGGAAAT CTCCAGTTTC GTCAGCAAAT 16560

CAATAGTAAG TTAGAATACC AAAAAGCTTT TGAGATTGCT TGCAAGATAG AGGGCTATCC 16620

AAGGCAAACC TCTGTCCATG CGGCTGGTGT TGTAATTAGT GACCAAGATT TAACCAACTA 16680

CATTCCTCTA AAGTATGGTG ATGAAATTCC ACTGACTCAG TATGATGCTC ATGGAGTTGA 16740

GGCTAGCGGA CTTTTGAAGA TGGACTTTCT GGGACTACGA AATTTGACCT TTGTCCAGAA 16800

GATGCAAGAG TTGCTTGCTG AAACAGAAGG TATTCATCTG AAAATTGAAG AAATCGATTT 16860

AGAAGACAAA GAAACGTTAG CTTTATTTGC CTCTGGTAAT ACAAAAGGTA TCTTTCAATT 16920

TGAGCAACCA GGTGCCATTC GTCTGCTTAA GCGTGTGCAA CCAGTCTGTT TTGAAGATGT 16980

CGTCGCGACT ACTTCTCTAA ATCGACCGGG TGCTAGTGAC TATATCAATA ATTTTGTGGC 17040

AAGAAAGCAT GGGCAGGAAG AAGTGACTGT TCTGGATCCA GTACTGGAGG ATATTTTGGC 17100

TCCAACCTAC GGCATAATGC TCTATCAGGA GCAGGTTATG CAGGTTGCCC AGCGACTTGC 17160

CGGATTTAGT CTTGGGAAAG CCGATATTTT GCGTCGGGCT ATGGGGAAAA AGGATGCCTC 17220

TGCCATGCAT GAGATGAGGG CTTCCTTTAT TCAAGGTTCA TTAGAAGCTG GTCATACTGT 17280

GGAAAAAGCA GAGCAGGTCT TTGATGTTAT GGAGAAGTTT GCAGGTTATG GTTTTAACAG 17340

GTCACACGCC TATGCCTACT CAGCCTTGGC CTTCCAGTTG GCTTATTTCA AAACGCATTA 17400

TCCAGCCATT TTTTATCAGG TCATGTTAAA TTCTTCCAAC AGTGATTACT TAATAGATGC 17460

ACTTGAAGCA GGTTTTGAAG TAGCCTCTCT ATCCATCAAC ACCATTCCCT ATCACGATAA 17520

AATTGCCAAC AAGGCCATCT ATCTAGGTTT GAAATCCATT AAAGGAGTCA GTAATGATTT 17580

AGCTCTCTGG ATTATTGAAA ATAGACCTTA TTCTAACATT GAAGATTTTA TAGCTAAATT 17640

ACCTGAGAAT TATCTGAAAC TTCCTCTGCT AGAACCTTTG GTAAAAGTTG GTCTTTTCGA 17700

TTCATTTGAA AAAAATCGTC AAAAAGTATT TAATAACTTA GCTAATCTAT TTGAATTTGT 17760

GAAAGAGTTG GGAAGTTTGT TTGGAGATGC TATTTATAGT TGGCAGGAAT CGGAAGATTG 17820

GACGGAACAA GAAAAATTTT ATATGGAACA AGAGCTTTTA GGGATAGGTG TCAGCAAACA 17880

TCCACTACAA GCTATTGCAA GTAAGGCTAT TTACCCGATT ACCCCAATCG GAAATTTGTC 17940

AGAAAATAGC TATGCTATTA TCTTGGTTGA AGTTCAGAAA ATAAAAGTGA TTCGTACCAA 18000

AAAGGGTGAA AATATGGCCT TCTTACAGGC AGATGATAGT AAGAAAAAAT TGGATGTCAC 18060

TCTCTTTTCA GACTTATATC GTCAGGTTGG ACAGGAAATA AAAGAGGGAG CCTTCTACTA 18120

TGTAAAAGGA AAAATACAAT CACGTGATGG CCGTCTGCAA ATGATTGCAC AAGAAATAAG 18180

AGAAGCAGTT GCTGAACGCT TTTGGATACA GGTGAAAAAT CATGAATCGG ATCAAGAAAT 18240

TTCACGCATT TTAGAACAAT TTAAAGGCCC AATCCCAGTC ATCATCCGGT ATGAAGAGGA 18300

ACAGAAAACC ATCGTTTCTC CCCATCATTT TGTAGCTAAA TCCAATGAAT TAGAGGAGAA 18360

ATTGAATGAA ATCGTTATGA AAACGATTTA TCGCTAAAAA TACGGAAAAT AGAAGAATTT 18420

TCAACGTAAA TGTGGTATAA TCAGTAAGAA TGTTAAAAGA AAAAGGAGCA TAACCAATAT 18480

GAAACGTATT GCTGTTTTGA CTAGTGGTGG AGACGCCCCT GGTATGAACG CTGCCATCCG 18540

TGCAGTTGTT CGTCAAGCAA TTTCAGAAGG AATGGAAGTT TTTGGTATCT ATGACGGATA 18600

TGCTGGTATG GTTGCCGGTG AAATTCATCC CCTAGATGCA GCTTCAGTAG GGGACATCAT 18660

TTCTCGTGGT GGTACTTTCC TTCACTCAGC TCGTTACCCA GAGTTCGCTC AACTTGAAGG 18720

GCAACTTAAA GGGATTGAGC AATTGAAAAA ACACGGAATT GAAGGTGTAG TTGTTATCGG 18780

TGGTGACGGA TCTTACCACG GCGCTATGCG TTTGACTGAA CATGGCTTCC CAGCTATTGG 18840

TCTTCCAGGT ACAATCGATA ACGATATCGT TGGTACTGAC TTTAC, TCG GTTTTGACAC 18900

AGCGGTTACT ACTGCCATGG ACGCTATCGA TAAGATTCGT GATACATCAT CAAGTCACCG 18960

TCGTACTTTT GTAATCGAAG TTATGGGACG TAACGCTGGT GATATCGCTC TTTGGGCTGG 19020

TATTGCAACT GGTGCTGATG AAATCATCAT CCCTGAAGCA GGCTTCAAGA TGGAAGATAT 19080

CGTAGCAAGC ATCAAAGCTG GTTATGAATG TGGTAAAAAA CACAATATTA TCGTCTTAGC 19140

TGAAGGTGTG ATGTCAGCGG CTGAATTTGG TCAAAAACTT AAAGAAGCTG GAGATACAAG 19200

CGACCTTCGT GTAACAGAAC TTGGACATAT TCAACGTGGT GGTTCTCCAA CTGCGCGTGA 19260

CCGTGTTTTG GCGTCACGTA TGGGTGCACA TGCTGTTAAA CTTCTTAAAG AAGGTATCGG 19320

TGGTGTTGCG GTTGGTATTC GTAACGAAAA AATGGTTGAA AATCCAATTC TTGGTACTGC 19380

AGAAGAAGGG GCATTGTTTA GCCTTACTGC AGAAGGTAAG ATTGTGGTTA ACAACCCAGC 19440

TACAAA 19446

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16593 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

TCGTAAATAT GCTCTGTTTT TGGATTTTGT TTCTTAATCT GTTTGGCAAG TGCCTTCATC 60

ATAGAAATAG GACCACACAT ATAGACGGTT GCATGTTCGG GCACTTCTTT TTGTTCAAAA 120

TTAAGATAGC CGTCTTTCGT ACTGTCGATT AGATGGAGTT CAAAATTAGG ATTTTTCTGA 180

GCATAGTTAC GGAGTAAATC TAGGTAGACT GCATTTTCAT CTCCACGGAA GCTATAGTAG 240

AAGTGAACCT GTTTATCTAA AATAGGATGT TCACGGATGT AAGAGATGAA GGGGGTGATC 300

CCAATACCTC CAGCAATCCA AACCTGATTT TCTCGTCCTT CTTCTATGAT CATGTGTCCG 360

TAAGCTCTGT CTAGGGTTAC TTTGCTGCCG GCTTGAAGAT TATCATAGAT ATTCTTGGTA 420

TGGTCGCCTG AAGTTTTAAC AGTAAAGTAA AGAGTTTGAC CATGACCTCC TGAGATAGAA 480

AAGGGATGCG GAGCACTTTC AAAGCCTTCT TGGAAAATCT TTAGAAAGGC AAATTGTCCT 540

GATTGATAGT TGAAAGGTCT GCTAAGATGG ATTTGAATTT CTCTAGTATC GTGATTTAAG 600

CGTTTGAGAT GGGTAATTTT CCCTAGATAG GGGAAGGAAA TCTTTTGATA TAGAAAAATG 660

ATATAAAAAC CAGCTAGTAA GCCTAAAAGG GCATAGCTAC CAACAAGAAA ACTTAGAAGA 720

TTAAATGTAA GGAGACGATT GCCCATTATC ATGTAGATGT GAAAGAGTCC TAAAATATAG 780

GCTAGGTAAA CCAGGCGGTG AATCCATCGC CAAGCTTCGT ATTGGATGTA TTTGCCTAAA 840

TAGGCGACAA GGATGATGCT GGCAAAGATA TAGATGGCAA GATTGCCAAA CTGAGCAGCT 900

AAGCGAGAGC CCCACAAACC GCCCATACTA AAGTTATGAA AGATTAGTAG GATGATTGAG 960

AGAAAGGCTG TGAATTTGTG GACGGTGTAG ACCTTCTCCA AACTGTGAAA CCAGCTTTCT 1020

AGTAGTGGGA GACGAGTGGC TAGGATAAAA GTCAGAGATA GGCTTGTTAA AGCTAGTCCT 1080

GGAATCATGA ATTGGGGAGA AGTGTTCATC CAAGTCAAAA GAGTCAAGAT AAAACTAGCT 1140

ATGATAAAGA GTAGTCCTTT GACTGATTTC ATAGAAAATT CCATTTCATT TAGATTTCGA 1200

TTTGTTGTAA ATAAATTTGT TACATTTTAT CATAGAAAAT GTATGGTGTC AAATTGAGGT 1260

CTATAAATAT CTACTCTCAT CAAAAAACTC TCCAATTGAA CTGGAGAGTG GCTGTTTATA 1320

CTCAATGAAA ATCAAAGAGC AAACTAGGAA GCTAGCCGCA AGTTGCTCAA AACACTGTTT 1380

TGAGGTTGCA GATAGAGCTG ACGTGGTTTG AAGAGATTTT CGAAGAGTGT TATTCTGCAG 1440

CTTGTTGCCA ACGTTTGGCT AGCATATGAG ACAGGCTAGA AATTGCTAGG TTAAAGCTGA 1500

AGTAGATGAG GGCAATCAGG ATGTAAAGAC TGAAGACCTG CTCTGGTTCG AAATAACGGC 1560

CCATGAGAAT TTGGCTGGCT CCAAAGAGTT CTTGTAGGGC GATAACAGAG TAGAGGAGAC 1620

TGGTATCCTT AATCACGGTA ACAAACTGAG AAATGATGGC TGGTAGCATT TTGCGGATGG 1680

CTTGTGGGAG AATGATGTAG TAGAGGATTT GGGCTGAGGT GAAGCCTTGT GACATTCCTG 1740

CTTCGTACTG TCCCTTGTCT ACGGCATTGA GACCGCCTCG AATAATCTCA GCCAAGGCTG 1800

CTGATGTAAA GAGAGTAAAG GCTGTAATAC CTGCTGGTGT GGATTTCATT TTGAACACCA 1860

AAAAGATAGT AAAAATCCAG AGAAGGTTGG GAACGTTGCG CACAAACTCG ATATAAATAC 1920

TGGAAATAAT GCGTAAGACA GGATTTTTGC CATTTCTCGT GACAGCTAGC ACCGTACCGA 1980

TGATAGTAGA GAGGATGATG GCAATCAGAG AAATATAGAG GGTCAAGCCA AATCCTTTAA 2040

AGATAAAGAC TAGGTTATCT GGGGTTAAAA CTTCTAAAAT AGATTCCATA GTAACCTCCT 2100

AAAGTGAATA GGCTTTTTTG TTGGCTTGCT CCATCTTGCG ACCAAACTGG GCAACAGGGA 2160

AGCATAGAGC AAAGTAGAGA AGAGCAGCAC CTAAAAAGGC TGGTATATAG TTTCCGTTGA 2220

GAGCCGACCA AGACTTAGTC ACAAACATCA AGTCTACTCC AGAGATGATA GCTACAGTAG 2280

AGGTGTTCTT GATGAGGTTA ACAATTTGGT TGGTCAATGG AGGGAGAATG ATGCGGAAGG 2340

CCTGAGGCAA GATAATCAAG CGCATGGCAC TGATATAGGT AAAACCTTGC GACAAGGCGG 2400

CCTCCATCTG ACCACTAGGA ATAGACTGAA TCCCTGAACG AATAACCTCA GCGATATAAG 2460

CGCCGTGATA GAGTCCCACG CAGAGAACGG CTGTCCAATA AATTGGAATC ATGATGATAT 2520

GGTCACTGAT AAGAGGTAGG CCATAAAAAA CAATAACAAA CTGCACCAAG AGGGGAGTAT 2580

TTTGGTAAAA TTCAACAAAG ATGCGAGCTA AAATGCGTAA AATTGGACGT TTACTGGTTG 2640

ACATGGCACC AAAGAAGATG CCCAAAACCA TAGCGAGGAT AAAGGAACCA ACCGCTAGGG 2700

CAAGGGTGAA GAGGAAACCA TTGAAAAATT GTCCAAAATC CTGAAAATAG GCTGTCCAAG 2760

ATGATAAATC TGTCATGGGG TGTCCTCCTT AATCTGCAGT ATGGCTAGAT GGTTTGAGCT 2820

TGTAACGGTC ATAAAGTTTC TGCAAACTAC CATCCTTGCT CCATTTAGTA ACCAAGTTAT 2880

CAAGATAGTC GTTGAGCTCT GTATTTGATT TCTTGGTAAC AATACCGTAG TCAGATGGCT 2940

TGAAACTATC ATCTAGTAGT GCTGTCCGTT TACTAGTGTA GCCAGATAGA ATAGAGCGGT 3000

CAACGGAAAA GGTATCGATA CGATGAGCGT GCAGGGAAGT AATCAATTCT GGGTAGGAAC 3060

CAAGTTCGAC GAATTTAAAC TTCAGACCTT TCTTTTTACC CAGTTCAGTA ATCAGGCGTT 3120

GGGTGATAGA ACCTTGGGCG ACTCCGATGG TTTTGCCGTT TAGGTCCTCA ATCTTTTTGA 3180

TTTTGGCAGA TTTATTGACC AAAAATCCAG AAGCGTCTGT GTAGTAGGGA CTGGTAAAGT 3240

TGTAGAGTTT TTTGCGTTCG TCCGTGATGG TAAAGGTCGC GATATCCATA TCGACCTGTT 3300

CATTGTCTAG AAGGGGGCCG CGGGTTTGTG CTGTAACCGG CACATAGCGA ATCTTGACCT 3360

TGAGTTCATC AGCTACCATC TTGGCCAAGT CGGTTTCGAT ACCAGAATAA GTACCGGTCT 3420

TGGGATCTTT GTAACCAAAA TTGGGAACGT CTTGTTTGAC ACCGACAACC AGTTCGCCTC 3480

TTTTTTGAAT GTCTGCGATA CTTGTATCAG CCTGGACTGG TTTGGCAGCA GCAAGGCCGA 3540

AAAGGCTAAT CAATAATGCT GATAAAAAGA ATTTTTTTTC ATAGGCGCCT CCTTATTTGA 3600

CTTTGTCACT TTCGTGGTTG ATAATTTTGC TGAGGAATTG TTGGGCACGA GGTTCGCTTG 3660

GATTGTCAAA AAAGTTATCG ACATCTGTCG TATCTACTAA AACTTCTCCG TCGGCCATAA 3720

AGATAATGCG GTCCGCAACC TCTCGAGCAA AGCCCATTTC GTGGGTAACG ATGATCATGT 3780

TCATCCCATC ATGCGCCAGT TTCTGCATAA CTGCTAGAAC ATCTCCGATA GTCTCAGGAT 3840

CAAGAGCAGA TGTTGGTTCA TCAAAGAGGA GGAGTTCCGG ATGCATAGCA AGACCACGAG 3900

CGATGGCGAT CCGCTGTTTT TGTCCACCAG ATAGCATGGC GGGATAGGAA TCTTTCTTGT 3960

CCCACATATT TACAAATTCC AGATATTTTT GGGCGGTTTT TTCAGCTTCT TTTTTATCAA 4020

TTCCTAGAAC TTCAATGGGT GCAAGCGTTA CGTTTTCTAA CACAGCTTTG TGTGGATAAA 4080

GGTTAAAATG TTGAAAAACC ATGCCGACTT CCTTGCGAAG AGGTACCAAA TCTTTCTGGC 4140

TGGCACCAGC AACTTGGTGC CCATTGACTA GGAGACTTCC TTTGTCAACA GTCTCTAAAC 4200

CATTGATCGT ACGGATAAGA GTGGACTTCC CAGAGCCAGA AGGTCCAAGC AGGACAACAA 4260

CTTGTCCTTT TTCAAAACGG AGATTGATGT TGCGGAATGC GTGGTAGTCT CCGTAATATT 4320

TTTCGACGTT TTTAAATTCT ACTAAAGCCA TGAGAGATCT CTATTGTGTT ATATTTTATA 4380

ACACGGTTCT ACAATAAAAG AATGTTCTTG TCAAATCATA TCTGAAAAAA TTCACTATAG 4440

TGAAATAAGA ACAGGAAAAA TCGATCGGGA CAGTCAAATC GATTTCTAAC AATATTTTAG 4500

AAGTAGAGGT GTACTATTCT AGTTTCAATA TACTATAAAA TGTTATAAAA AAGCAATCTG 4560

GATAGAGAAA ACGTCTAAAT CATGTTATAA TGAAGCAATA GAATTCTTAG AAAGAGTGGA 4620

TGTCTTTTTG ATAACACCTA CTTATGAATG GCAGTTTGCC CTGCAGGTAG AAGATGCGGA 4680

TTTTACAAAG ATAGCCAAGA AGGCTGGACT GGGTCCTGAG GTGGCTCGGT TATTGTTTGA 4740

GAGAGGGATT CAGAACCAAG AAAGTCTGAA GAAGTTTTTA GAACCTTCCT TGGAGGACTT 4800

ACATGATGCT TATCTGCTCC ATGATATGGA CAAGGCAGTG GAGCGGATTC GTCAGGCTAT 4860

TGAAGAAGGG GAAAATATTC TTGTTTATGG AGACTATGAT GCGGATGGCA TGACTTCGGC 4920

TTCTATTGTG AAGGAAAGTT TGGAACAACT TGGTGCTGAG TGCCGAGTTT ACCTGCCAAA 4980

TCGTTTTACC GATGGCTATG GCCCTAATGC TAGTGTTTAT AAATACTTTA TCGAGCAAGA 5040

AGGGATTTCC TTGATTGTGA CGGTGGACAA TGGGGTTGCT GGTCATGAGG CTATTGCATT 5100

GGCTCAGTCT ATGGGAGTAG ATGTCATTGT GACAGACCAT CATTCCATGC CTGAAACCCT 5160

GCCAGATGCT TATGCTATTG TCCATCCTGA ACATCCAGAT GCGGATTATC CTTTTAAATA 5220

TTTGGCTGGT TGTGGAGTTG CTTTCAAGTT GGCTTGTGCC CTGTTAGAAG AAGTGCAAGT 5280

GGAATTGCTT GATTTGGTCG CTATTGGAAC TATTGCAGAT ATGGTGAGTC TGACGGATGA 5340

AAATCGTATC TTAGTTCAAT ATGGTCTGGA AATGTTGGGT CATACCCAGC GCATTGGTCT 5400

GCAAGAAATG CTGGACATGG CTGGGATTGC TGCCAACGAA GTAACAGAAG AAACGGTTGG 5460

TTTCCAGATT GCTCCTCGTT TGAATGCCTT GGGTCGCTTG GATGATCCCA ATCCTGCCAT 5520

TGATTTGTTG ACTGGATTTG ATGATGAGGA AGCGCATGAG ATTGCCCTTA TGATTCACCA 5580

GAAAAACGAA GAGCGCAAGG AAATCGTTCA GTCTATCTAT GAAGAAGCCA AGACCATCGT 5640

GGATCCTGAG AAGAAGGTTC AGGTCTTGGC CAAGGAAGGC TGGAATCCTG GGGTTCTAGG 5700

AATCGTGGCT GGTCGTTTAT TGGAAGAATT GGGACAGACA GTCATTGTTC TTAATATAGA 5760

AGACGGTCGT GCCAAGGGCA GTGCTCGTAG TGTGGAAGCG GTCGATATTT TTGAAGCTCT 5820

GGATCCCCAT CGAGACCTCT TCATCGCCTT TGGAGGTCAT GCAGGTGCAG CGGGTATGAC 5880

GCTGGAAGTT GAGCAACTCT CAGATTTATC TCAGGTTTTG GAAGATTATG TTCGTGAAAA 5940

AGGTGCAGAT GCTGGTGGCA AGAATAAGTT AAACCTAGAT GAAGAGTTGG ATTTGGAGGC 6000

ACTTAGCTTG GAAACGGTCA AAAGTTTTGA ACGTTTAGCT CCTTTTGGAA TGGATAATCA 6060

GAAACCTATT TTTTATATCA AGAATTTTCA GGTCGAAAGT GCTCGTACTA TGGGGGCAGG 6120

TAATGCCCAT CTAAAGCTGA AAATTTCCAA GGGTGAGGCG AGTTTTGAAG TGGTAGCCTT 6180

TGGTCAAGGC AGATGGGCGA CAGAGTTTTC TCAAACCAAG AATCTAGAGT TAGCGGTTAA 6240

ATTGTCTGTC AACCAATGGA ATGGCCAAAC TGCCCTCCAG TTGATGATGG TGGATGCGCG 6300

AGTGGAAGGT GTTCAACTTT TTAACATTCG TGGAAAAAAT GCAGTCTTGC CAGAAGGTGT 6360

TCCAGTCTTG GATTTTCCTG GAGAACTGCC AAATCTTGCG GCTAGTGAAG CTGTTGTCGT 6420

AAAAAACATT CCAGAGGATA TTACTCAGCT GAAGACCATT TTTCAGGAAC AGCATTTCTC 6480

TGCTGTCTAT TTCAAAAATG ATATTGACAA GGCTTATTAT CTGACAGGTT ATGGGACTAG 6540

AGATCAGTTT GCCAAATTGT ACAAGACTAT TTACCAGTTC CCAGAGTTTG ATATTCGCTA 6600

CAAGCTGAAA GATTTGGCTG CATATCTTAA TATTCAACAA ATCTTGCTGG TCAAGATGAT 6660

TCAAGTATTT GAAGAACTAG GCTTTGTGAC GATAAAAGAT GGTGTGATGA CAGTCAATAA 6720

AGAGGCGCCA AAGCGGGAGA TAGGAGAAAG TCAAATTTAC CAAAATCTCA AACAAACCGT 6780

TAAAGACCAA GAAATGATGG CGCTGGGTAC GGTGCAAGAA ATTTATGATT TTTTGATGGA 6840

AAAAGAGTAG AAGTTAGGAA AGAGTTGGGA AATCAACTCT TTTTTGAAAA CAGACCTTCA 6900

TTTTGAAAAT CATCAAAAAA ATGGTATAAT GGTAGGAAAA GATTCGGCTG AAAGTATCAG 6960

AACTTTTAGA ATAAGAGGGT AGAATTGCCC TATAATCAAG ATAAACTAAG ATTTTGGAGG 7020

AAAAATGAGT AATATCAGTT TAACAACACT TGGTGGTGTG CGTGAGAATG GAAAAAATAT 7080

GTACATTGCT GAAATTGGAG AGTCCATTTT TGTTTTGAAT GTAGGGTTAA AATATCCTGA 7140

AAATGAACAA TTAGGGGTCG ATGTGGTGAT TCCAAACATG GATTACCTTT TTGAAAATAG 7200

CGACCGTATT GCTGGGGTTT TCTTGACCCA CGGGCATGCG GATGCCATTG GTGCTCTACC 7260

GTATCTCTTG GCAGAGGCTA AAGTTCCTGT ATTTGGGTCT GAGTTGACCA TTGAGTTGGC 7320

AAAGCTCTTT GTCAAAGGAA ATGATGCCGT TAAGAAATTT AATGATTTCC ATGTCATTGA 7380

TGAGAATACG GAGATTGATT TTGGTGGGAC AGTGGTTTCC TTCTTCCCTA CGACTTACTC 7440

CGTTCCAGAG AGTCTGGGAA TTGTCTTGAA GACATCGGAA GGAAGCATCG TTTATACAGG 7500

TGACTTCAAA TTTGACCAAA CGGCTAGTGA ATCTTATGCA ACTGATTTTG CTCGTTTGGC 7560

AGAGATTGGT CGTGACGGCG TCCTGGCTCT CCTCAGTGAT TCGGCCAATG CAGACAGCAA 7620

TATTCAGGTG GCTAGTGAAA GTGAAGTTAG GGATGAAATT ACCCAAACTA TTGCTGACTG 7680

GGAAGGTCGT ATCATCGTTG CAGCTGTTTC CAGTAATCTT TCTCGTATTC AGCAGATTTT 7740

TGACGCTGCG GATAAAACAG GTCGACGTAT CGTCTTGACA GGATTTGATA TTGAAAATAT 7800

CGTCCGCACA GCGATTCGTC TTAAGAAGTT GTCTTTAGCC AACGAAATTC TTTTGATTAA 7860

GCCTAAAGAT ATGTCTCGCT TTGAAGACCA TGAGTTGATT ATTCTTGAGA CAGGTCGTAT 7920

GGGTGAGCCT ATCAATGGAC TTCGTAAGAT GTCGATTGGT CGCCATCGTT ATGTAGAAAT 7980

CAAGGATGGG GACCTAGTCT ATATTGCTAC GGCTCCGTCT ATTGCTAAAG AAGCCTTTGT 8040

TGCGCGTGTG GAAAATATGA TTTATCAGGC AGGTGGGGTT GTCAAATTGA TTACCCAAAG 8100

TTTACATGTA TCAGGGCACG GAAATGTGCG TGATTTGCAG CTGATGATCA ATCTTTTGCA 8160

ACCTAAGTAC CTCTTCCCTG TCCAAGGGGA GTATCGTGAG TTGGATGCTC ACGCTAAGGC 8220

TGCCATGGCA GTTGGGATGT TGCCAGAACG CATCTTCATT CCTAAAAAGG GGACGACCAT 8280

GGCTTACGAG AATGGAGACT TTGTTCCAGC TGGATCGGTT TCAGCAGGAG ATATCTTGAT 8340

TGATGGGAAT GCCATTGGTG ATGTTGGAAA TGTTGTTCTT CGTGACCGTA AGGTCTTGTC 8400

AGAGGATGGA ATTTTCATCG TGGCTATTAC AGTCAACCGT CGTGAGAAGA AAATTGTGGC 8460

TAGGGCTCGT GTTCACACGC GTGGATTTGT TTATCTCAAG AAGAGTCGCG ATATTCTCCG 8520

TGAAAGTTCA GAATTGATTA ACCAAACGGT AGAAGAGTAT CTTCAAGGAG ATGACTTTGA 8580

CTGGGCAGAT CTCAAAGGTA AGGTTCGTGA CAATCTGACC AAGTACCTCT TTGATCAAAC 8640

CAAGCGTCGC CCAGCCATTT TACCAGTAGT CATGGAAGCA AAATAATCGT TGAAATAAAC 8700

AGAGAGAAAG TCGAGTTTCG GCTTTTTCTT ATAGAAAAAT AGAAGGAGAA AATCATGGCA 8760

GTGATGAAAA TCGAGTATTA CTCACAAGTA TTGGATATGG AGTGGGGGGT GAATGTCCTC 8820

TACCCTGATG CCAATCGAGT GGAAGAACCA GAGTGTGAAG ATATTCCCGT CTTGTACCTT 8880

TTGCACGGGA TGTCTGGAAA TCATAATAGT TGGCTTAAGC GGACCAATGT AGAACGCTTG 8940

CTTCGAGGAA CTAATCTCAT CGTTGTTATG CCCAATACCA GCAATGGTTG GTACACCGAT 9000

ACCCAGTATG GTTTTGACTA CTACACGGCT CTAGCAGAGG AATTGCCACA GGTTCTGAAA 9060

CGCTTCTTCC CTAATATGAC GAGCAAGCGT GAAAAGACCT TTATCGCTGG TCTTTCTATG 9120

GGAGGCTACG GCTGCTTCAA ACTGGCTCTT ACGACAAATC GTTTTTCTCA TGCAGCTAGT 9180

TTTTCAGGTG CCCTCAGCTT TCAAAACTTT TCTCCTGAAA GTCAAAATCT GGGAAGTCCA 9240

GCCTACTGGA GAGGTGTTTT TGGAGAGATT AGAGACTGGA CAACTAGTCC CTATTCTCTT 9300

GAAAGTCTGG CTAAAAAATC GGATAAAAAG ACCAAACTTT GGGCGTGGTG TGGCGAACAG 9360

GATTTCTTGT ACGAAGCCAA TAATCTCGCA GTGAAAAATC TCAAAAAACT AGGTTTTGAT 9420

GTGACCTATA GCCATAGCGC TGGAACTCAC GAGTGGTACT ACTGGGAAAA ACAATTGGAA 9480

GTTTTTTTAA CAACCCTACC AATTGATTTC AAATTAGAAG AGAGACTGAC TTAGTTTGAA 9540

CTTCAGCATA GGGGGAGTAG AACTAAAATA AAATATGTTT TCACTAGACT TTTCAAACGm 9600

AAGTAGTAGA ATAGTAATAA AATACTGGAG GAAAGAGAGT AGGAAATGTA CCGTTATCAA 9660

ATTGGCATTC CCACATTAGA ATATGATCAG TTTGTCAAAG AACATGAATT AGCCAATGTA 9720

TTACAAAGTA GTGCTTGGGA GGAAGTTAAG TCTAATTGGC AACATGAGAA GTTTGGTGTT 9780

TACAGGGAAG AAAAATTACT GGCGACAGCT AGTATTTTGA TTAGAACTCT TCCGCTAGGC 9840

TATAAAATGT TTTACATCCC AAGAGGACCT ATATTGGATT ATGGGGATAA AGAACTCTTG 9900

AATTTTGCCA TTCAGTCTAT TAAGTCCTAT GCTCGCAGTA AGAGAGCGGT TTTTGTGACT 9960

TTTGACCCAA GTATTTGCCT ATCTCAAAGT TTAATCAATC AGGAAAAGAC AGAATTTCCT 10020

GAAAATCTGG CTATTATTGA TAGTTTGCAA CAAATGGGAG TAAGGTGGTC AGGAAAAACG 10080

GAGGAAATGG GAGACACCAT TCAACCTCGT ATTCAGGCGA AAATATACAA GGAAAATTTT 10140

GAAGAAGATA AACTTTCCAA GTCAACAAAA CAGGCTATTC GAACAGCACG AAACAAAGGG 10200

CTTGAGATTC AATATGGTGG ACTGGAACTA TTAGATTCAT TTTCGGAGTT GATGAAAAAA 10260

ACTGAGAAGC GAAAAGAGAT TCATTTGAGG AATGAAGCCT ATTATAAAAA ATTGTTAGAT 10320

AATTTTAAGG ACAAGGCCTA TATCACCTTG GCCACCTTGG ATGTTTCTAA ACGTTCGCAA 10380

GAGTTAGAAG AACAGTTAGC GAAAAATAGA GCCTTGGAAG AGACCTTTAC TGAGTCGACT 10440

CGAACTTCAA AAGTAGAAGC GCAGAAGAAG GAAAAAGAAC GTTTGTTAGA GGAATTGACC 10500

TTCTTGCAGG AATATATAGA TGTAGGTCAA GCGAGAGTTC CTTTAGCGGC TACTTTGAGT 10560

TTGGAATTTG GTACTACCTC TGTCAATATA TATGCTGGTA TGGATGATGA TTTTAAACGT 10620

TACAATGCAC CAATTTTAAC TTGGTATGAA ACGGCTCGCT ATGCCTTTGA ACGAGGTATG 10680

ATCTGGCAAA ATTTAGGTGG TGTTGAAAAC TCTCTCAATG GTGGACTTTA TCATTTTAAG 10740

GAAAAATTTA ATCCAACGAT TGAAGAATAC TTGGGTGAAT TTACAATGCC CACTCATCCT 10800

CTCTATCCTC TGTTAAGACT TGCTCTTGAT TTCCGTAAAA CATTAAGAAA AAAACATAGA 10860

AAGTAAGTAT ATGGCACTAA CAACACTCAC GAAAGAAGAG TTTCAGACTT ATTCTGATCA 10920

GGTTTCTTCT CGTTCCTTTA TGCAATCTGT CCAGATGGGG GATTTGCTAG AAAAAAGAGG 10980

GGCTCGAATT GTTTATCTTG CTTTGAAACA AGAAGGAGAA ATTCAAGTTG CAGCTCTGGT 11040

TTATAGCCTG CCCATGCTGG GTGGTCTGCA TATGGAACTC AATTCGGGGC CGATTTATAC 11100

CCAACAAGAT GCTCTTCCAG TTTTTTATGC AGAGTTAAAA GAATATGCCA AGCAAAATGG 11160

TGTATTAGAG TTGCTTGTAA AACCCTATGA AACTTATCAA ACTTTTGATA GCCAAGGTAA 11220

TCCAATAGAT GCTGAGAAAA AAAGTATTAT TCAAGATTTG ACTGATTTAG GTTATCAATT 11280

TGATGGCTTA ACAACAGGTT ACCCAGGTGG AGAACCAGAT TGGTTATACT ATAAAGATTT 11340

AACTGAATTA ACTGAAAAGA GTTTGCTTAA AAGTTTTAGC AAAAAGGGTA AACCCTTGGT 11400

GAAAAAGGCT GAAACCTTTG GCATTCGGTT GAAAAAGTTA AAACGTGAAG AACTATCGAT 11460

TTTTAAGAAT ATAACAAAAG AAACCTCTGA ACGTAGAGAA TATAGTGATA AAAGTTTAGA 11520

ATATTATGAG CATTTTTATG ATACTTTTGG AGAACAAGCG GAGTTTCTCA TAGCAAGCTT 11580

AAATTTTTCG GACTATATGA GCAAATTGCA AGGTGAACAA AGTAAACTAG AAGAAAACTT 11640

GGACAAGTTG CGACTTGATT TGAGTAAAAA TCCTCATTCT GAGAAAAAAC AAAATCAACT 11700

GAGAGAATAT TCTAGTCAAT TTGAAACGTT TGAAGTTCGA AAAGCAGAAG CGCGAGACTT 11760

GATTGAAAAA TATGGAGAAG AAGATATTGT TTTAGCTGGG AGTTTATTTG TTTATATGCC 11820

TCAGGAAACG ACTTATCTCT TTAGTGGTTC CTACACTGAG TTTAATAAGT TCTATGCCCC 11880

TGCACTGCTT CAAAAATATG TTATGTTGGA AAGCATAAAA CGTGGAATAC CTAAATACAA 11940

CTTCCTAGGC ATTCAAGGGA TTTTTGATGG AAGTGATGGT GTTTTGCGTT TTAAACAGAA 12000

TTTTAATGGC TATATTGTAC GCAAAGCAGG TACTTTCCGT TACCATCCAT CGCCTTTAAA 12060

ATACAAAGCT ATCCAGTTAC TCAAAAAAAT AGTAGGACGT TAAGATGAAA AAGTCAGTAT 12120

TTAGATTTCT TTTAGCTTCT TTTAGTAAAA TAATTCTTAT TTGCTAGAAA GGTGGAGAGA 12180

CATGCGCTGG CTTTTTCGTT TGATAGGGGC TTTCTTTTCT TTTGTGTGGC GTTTGTTTTG 12240

GCGTCTGGTT TGGATAGTTG TGCTCTTATG TGTGCTTGCT TTCGGACTTC TCTGGTATCT 12300

GAACGGAGAT TTTCAAGGAG CGCTAAAGCA AGCAGAACGG TCAGTAAAAA TTGGTCAACA 12360

AAGTATTGAC CAATGGGAGA AAACAGGGCA ACTGCCTAAG TTAAGCCAGA CAGATAGTCA 12420

CCAGCATTCT GAAGGAAGGT GGGCACAGGC CTCTGCTCGT ATTTACCTGG ATCCGCAGAT 12480

GGATTCACGC TTTCAAGAGG CTTATTTAGA AGCAATCCAG AACTGGAATC AAACTGGTGC 12540

TTTTAACTTT GAACTCGTGA CTGAGTCTAG TAAGGCGGAT ATTACGGCTA CGGAGATGAA 12600

CGACGGAGGC ACTCCTGTGG CAGGAGAGGC GGAAAGTCAA ACTAATCTCT TAACAGGGCA 12660

ATTCTTGTCC GTAACGGTGC GGTTGAATCA TTATTATTTG TCCAATCCAT ACTATGGCTA 12720

CTCCTATGAA CGCCTTGTCC ATACGGCAGA ACATGAGTTA GGTCATGCGA TTGGCTTGGA 12780

CCATACAGAT GAGAAGTCTG TCATGCAACC AGCAGGTTCC TTTTATGGTA TCCAGGAAGA 12840

GGATGTTGCA AACCTCCGAA AAATATATGA GACTAGTGAG TAGGGTACTA TCTTTCCCTA 12900

CTTTTTTTGC TATAATGGAA CTATGAACAA CTTGATTAAA TCAAAACTAG AGCTCTTGCC 12960

GACCAGCCCT GGTTGCTACA TTCATAAGGA TAAAAATGGC ACCATTATCT ATGTAGGAAA 13020

GGCTAAAAAT CTGCGTAATC GAGTACGGTC CTATTTTCGT GGAAGTCATG ATACCAAGAC 13080

AGAGGCTCTG GTGTCTGAAA TTGTGGATTT TGAATTTATT GTTACGGAGT CTAATATTGA 13140

GGCACTTCTC CTAGAAATCA ACCTGATCAA GGAAAACAAG CCCAAGTACA ATATCATGCT 13200

CAAGGATGAC AAGTCCTATC CTTTCATCAA AATCACCAAT GAGCGCTATC CACGCTTGAT 13260

TATCACTCGT CAGGTCAAAA AGGACGGAGG TCTTTATTTT GGACCCTATC CCGATGTGGG 13320

GGCAGCCAAT GAAATCAAGC GGTTGCTGGA TCGGATATTC CCTTTTCGTA AGTGTACCAA 13380

CCCGCCCTCT AAGGTCTGTT TTTATTACCA TATCGGCCAG TGTATGGCCC ACACCATCTG 13440

TAAGAAGGAT GAGGCTTATT TCAAGTCTAT GGCCCAGGAG GTGTCTGATT TTCTGAAAGG 13500

TCAGGATGAC AAAATCATCG ATGATCTCAA GAGTAAAATG GCAGTAGCAG CACAAAGTAT 13560

GGAGTTTGAA CGTGCGGCGG AATACCGTGA CCTGATTCAG GCTATTGGAA CGCTTCGAAC 13620

CAAGCAACGG GTCATGGCGA AAGATTTGCA AAATCGCGAT GTCTTTGGCT ACTATGTGGA 13680

TAAGGGCTGG ATGTGTGTGC AGGTTTTCTT TGTCCGTCAG GtAAGCTCAT CGAGCGCGAT 13740

GTCAATCTCT TCCCCTACTT CAATGATCCA GATGAGGATT TTTTGACCTA TGTAGGACAA 13800

TTCTATCAAG AAAAATCTCA TCTAGTTCCC AATGAGGTAC TGATTCCGCA GATATTGACG 13860

AAGAAGCTGT CAAGGCTTTG GTGGATTCCA AGATTCTTAA GCCTCAACGT GGAGAGAAAA 13920

AACAACTGGT CAATCTAGCC ATAAAAAATG CTCGTGTTAG TCTAGAGCAG AAGTTCAATC 13980

TGCTAGAAAA ATCTGTCGAA AAGACTCAAG GAGCTATTGA AAATCTAGGG CGTTTGCTCC 14040

AAATCCCGAC CCCAGTACGT ATCGAGTCCT TCGATAACTC TAATATCATG GGAACTAGCC 14100

CTGTTTCGGC TATGGTGGTC TTTGTCAACG GTAAACCGAG TAAGAAGGAT TACCGTAAGT 14160

ACAAGATAAA AACGGTTGTT GGACCAGACG ACTATGCCAG CATGAGAGAG GTCATTCGCA 14220

GACGCTATGG TCGAGTACAG CGTGAGGCTT TGACTCCTCC AGATTTGATT GTGATTGATG 14280

GGGGGCAAGG TCAAGTCAAT ATCGCTAAGC AGGTTATCCA AGAGGAACTG GGCTTGGATA 14340

TTCCAATTGC TGGGCTGCAA AAGAATGATA AGCACCAAAC CCATGAATTG CTCTTTGGAG 14400

ATCCGCTTGA GGTGGTGGAT TTGTCTCGCA ATTCTCAGGA ATTTTTCCTC CTCCAACGCA 14460

TCCAAGATGA GGTGCACCGC TTTGCTATCA CTTTCCACCG CCAACTGCGC TCCAAAAATT 14520

CTTTCTCATC TCAATTGGAT GGGATTGACG GTCTGGGACC TAAACGCAAG CAGAATCTTA 14580

TGAAGCATTT CAAGTCTTTG ACCAAAATCA AGGAAGCCAG TGTGGATGAG ATTGTCGAAG 14640

TTGGGGTACC TAGAGTCGTT GCAGAGGCTG TGCAAAGAAA GTTGAACCCG CAGGGAGAAG 14700

CCTTGCCTCA AGTAGCAGAA GAAAGAGTAG ATTACCAAAC GGAAGGAAAC CACAATGAAC 14760

CATAAAATCG CAATTTTATC AGATGTTCAT GGCAATGCGA CGGCGCTAGA AGCAGTGATT 14820

GCAGATGCTA AAAATCAAGG GGCCAGTGAA TATTGGCTTC TGGGAGATAT TTTTCTTCCT 14880

GGTCCAGGCG CAAATGACTT AGTCGCCCTG CTAAAGGACC TTCCTATCAC AGCAAGTGTT 14940

CGAGGCAATT GGGATGATCG TGTCCTTGAG GCTTTAGATG GGCAATATGG CTTAGAAGAC 15000

CCACAGGAAG TTCAGCTCTT GCGTATGACA CAGTATTTGA TGGAGCGAAT GGATCCTGCA 15060

ACGATTGTCT GGCTACGAAG CTTGCCTTTG CTGGAAAAGA AAGAAATTGA CGGATTGCGC 15120

TTTTCTATCT CTCATAATTT ACCTGACAAA AACTATGGTG GTGACTTGCT AGTTGAGAAT 15180

GATACAGAGA AATTTGACCA ACTGCTAGAT GCGGAAACGG ACGTGGCAGT TTATGGTCAT 15240

GTTCACAAGC AGTTGCTTCG TTATGGAAGT CAAGGGCAAC AAATCATCAA TCCAGGGTCG 15300

ATTGGCATGC CCTATTTTAA TTGGGAGGCG TTAAAAAATC ACCGTTCCCA GTATGCCGTG 15360

ATAGAAGTTG AAGATGGGGA ATTACTCAAT ATCCAATTTC GTAAAGTTGC TTATGATTAC 15420

GAAGCTGAGT TAGAATTGGC CAAGTCCAAG GGGCTTCCCT TTATCGAAAT GTATGAAGAA 15480

CTGCGTCGTG ACGATAACTA TCAGGGGCAC AATCTGGAAT TATTAGCCAG CTTAATAGAA 15540

AAGCATGGGT ATGTAGAGGA TGTGAAGAAT TTTTTTGATT TTTTGTAAGA GTTTCCTAAA 15600

ATAGCCAATG CAAACTAAAA AAGCGATTTG CTGGTCCAAT CGCTTTTAGT ATATCTTATA 15660

CTCAATGAAA ATCAAAGAGC AAACTAGGAA GCTAGCCGTA GGTTGCTCAA AGCACAGCTT 15720

TGAGGTTGCA GATAAAGCTG ACGTGGTTTG AAGAGATTTT CGAAGAGTGT TATTGTAACT 15780

GAGATTGATC TGGGAGGTAA GAACCACCTA GATAGGTATT GCTGAGTTTT TCAAGGGTTC 15840

CGTCTTGATA GAGTTCTTTG AGCGCTTTAT CAAATTGCTC TTTAAACTCT TTTTGGTCGC 15900

TTGAGAAAAT GATATAATTG CTGGGGCTAT CTGCAGAAGG TAAATCAACG ACTGAGAGGT 15960

CTAAACCACG GTCCTTGATA ATCTTTTGAA CGGATACCTT GTCAAAAACT AGGAAATCAA 16020

ACTCTCCGTT AGCAAGGTCT AGGATTCGTT TACCAATATC CTCACCAGAA AAATTAATTG 16080

TAGCGGGATT ATCAGTGTGT TTCTGATTCC AGTTATTGAT GAATTGAGCG TTAGAAGTTC 16140

CGGTATCCTC TTGTGTTGTT TTACCAGCGA TCTGGTCAAG AGAAGTCAAA GGATTTTTCT 16200

TGTTGCTGAC AAGGACGAGG GGATTGTTGG AAATTGGAAG CGAGTAAAGG TATTTTTCAG 16260

CACGCTCTTT TGTGTAACTC AAGTTATTGG CCGCAGCCTG ATAGTGACCA GAATCAAGTC 16320

CTGGGAAGAT GCTCTCCCAG GCGGTTCTTT GGAATTGAAT CTCGTAGTCG CTGAGTTTTT 16380

CATCTACTGC CTTTAAAACT TCGATATCAA AGCCTGTCAG ATTGCCCTTG TCTTCGTAGT 16440

CAAATGGTGG CACGTCGCCA GCTGTAGCAA GGACGATTGT CTTTTGAGCG CTAGTCTCTT 16500

TGGGTGTAGC TTGATTCTCA CAGGCAACCA AAAATGGTAG GATAGCTAGT AATAGGCTAA 16560

ATTTTTTCAT ACTGTCTCCA TTCAAATGTA AAG 16593

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3510 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

GGGATATCCT TATATCCTTG TTCCTGGAAC CATTGTGGGA ATTGCTCAAC AGTTTTTTCA 60

CCTTGAATTC CTGGTGCAAT GACAGTAAGA ATTTCGAAAT CACGATCTGG TTTCGCCGCT 120

AGTTCCATCA ACTCTGGCAT ACTTTTCTTG CATGGACCAC ACCATGAAGC CCAAAACTTC 180

AAGTAAACCT TTTTACCCTT AAAATCAGAT AACTTAACTT CTTTGCCATC CATGGATTGC 240

AATGTGAAGT CTGGAGCATC TTTTCCAACA GCAATTTGTT GTACAGTCGT TTGTTGTTTT 300

GGCTGTTGTG CTGCTTGAGT CTTTTTAGTT TCTTCCTCAC CACAGGCCAT CAATACAACT 360

AATGACAAGA GACTTAAGCC AGCAAACATT ACTTTTTTCA TTTGTCCTCC TTTATTCAAA 420

AATTCCAGCT AGAACATTTA CTTGTCCTAA TAGTAACAAA ATTCCCATTA AAACAATGAG 480

GAAACCACCA ATTTTCTTTA GTAGCATCAT ATGACGCTTG ATTTTACTAA AATATGGCAT 540

GACTAGACCT GAAGCTAGTG CCAATACCAA GAAAGGAAGG GCCATGCCaG AGTGTAAATG 600

AGAGTATAAA TCGCTCCTTG CCAAGCGCCA TTGCCTCCAG AAGCCGCAAG TGCTAAAACA 660

GAACTTAAAA CTGGACCAAT ACAAGGTGTC CAACCAAAGC TAAAGGTAAT ACCAAGTAAA 720

AAAGCTGACC AATAACGATT AGAATCTGAT TTTTTAAAGG TAAAACTTTT TTGAACTTCT 780

AATTTCTTCA AATGAAAAAT TTCCATCTGG TGAAGACCCA AAATGATAAT AATAGCTCCC 840

ATGCCATATC GAAACCAATT TGCATAGAGA ATATGACCAA AGTAACCAGC ACCAAAGCCT 900

AGAATAAAGA AAATGAGAGA GATACCAGCG ATAAAGCAAA GTGTTCGAAT CAAGCCTGAC 960

CAGAGAACCT TTCTCCCAAA CAAAGAAAAG CTTTTTGCAC TTTCTTGATC ATCCAATAAA 1020

ATCCCAGCAT AGACTGGCAG AAGAGGAAAA ATACAAGGAG AAAAAAAGGA TAAAACACCT 1080

GCTAGAAAAA CAGAGATTAA AAATACTATC GTTTCCAATA AAGAACCAAC TTTCTTAATA 1140

ATTCTAATCC TATTTTACTA TATTCAATTT TATTTGTAAG CTTTCTGCTA CGCAAAATCG 1200

TATCGGGCAC TATTGGACCA ATCTTTTCTT TTGCTAGTCA AGGCGGATCT TATCCCCCAA 1260

AATAGCCAAA AAGCAACGAC AAGGATTACT CATCGCTGCT TTTGTGAACG AAAATGTCTT 1320

TTAGGTCTGA CATTTCATAA ATCATGTTTT ACTTGAGTTT GTCAAGGATT GCTTTAAGCT 1380

CCTCTACTAG TTTAGTTTCT GTCTCTGCTG AGCCATTTTC TTCTTTCACG AAATCAAGGG 1440

TTTCTTGGAG AAGGTTTTGG GCTTTGGCAA GGACTTTTTT ATCCGCTTTT TCTGCATCTA 1500

GCTGTCCTAG AACCTTGATC AATTCCGTGC TTAATTGCTG GATTT TGAC TCTTTCTTAC 1560

GGCGAATCAG CCAGAAGGCA ATCACGCCTA GGAGGGCAAG TAGACTGACC ACAATCACTC 1620

CTGCCGGAAC TGAGTTTGTT TCAGTCATCT TATCTGAATC CTTACTATCT TCCGTTCCTT 1680

GTTTTGCATC CTTCTTGTCC TGTGCAGGCT TGCTGTCGCT AGCATTTGCT TTCACATCTT 1740

TGAGAGAGTC CAAGGCAGCC CAGCCTTCAC AGACTCTACT GCAGTATGCA GACCTTACTC 1800

TGTCAAGGCA CTATCTTCCG GAGCTTTTTG AGCATCTAGG AGGACAGCCT TGGTTGCATC 1860

GATTTTCGGA TCAGATACTG TTGCCAAAGC TTTCAAGCGT TGGTCTAACT CTTGACTCAA 1920

GGCACGAAGT TCAGACTTGT CAACTTGCTC TTGAGCTTGT GTGCTCGTTG AGCTAGCCGA 1980

AGCGCTTGCT ACCACTCTAG GATCTTGAGT CGGAGCTGAG CTTGGAGCTG GGACAGGGCT 2040

TGCAGGTTGA CTAGGAACAG TTATGGTATA TTGAAACTAG AATAGTACAT ATGGACTTCT 2100

AAAACATTGT TAGAATTCGA TTTTACTGTC CTGATCGATT TGTCCTATTC TTATTTCATT 2160

TTACTATAAT AACCGATGGT GTGGTTAATG TTGGTAAGAG AAACTTCTGA AACCAAGCTT 2220

CAAAAAAGTC GCTCGTCATC GTCTCTTCGT AAGTCATTGG AGCGATTAAT TCACCATTTG 2280

TTAGACCTGC AACCAAAGAA ATCCTCTGAT ATCTTCTTCC AGATACTTTG CCTCTTATTA 2340

ACTGACCTTT TAATGAGCGA CCATATTCTC GATAAAAATA AGTATCGAAT CCTGTTTCGT 2400

CAATCTAAAC AGGTGCTAGG TGCTTTAAAC TATTAAAATT CTTAAGAAAT AAGGCTACTT 2460

TTTCTGGGTC TTGTTCATAG TAGGTGTGGT TCTTTTTTTC GAGTGTAGCC CATAGCTTTG 2520

AGCGCATAGT GGATGGTAGT TGGATGACAG CCAAAkTCAG AAGCTATTTC AGTCAAATAA 2580

GCrTCTGGAT TGTCAGTAAG ATAGTTTTTA AGTCTATCTC TATCAACTTT TCTTGGTTTT 2640

GTTCCTTTTA CTTGGTGGTT TAGCTCTCCT GTTTTCTCTT TTAGCTTTAA CCAGCCATAA 2700

ATGGTATTAC GTGAGATTTG GAAAACGTGT GATGCTTCTG TTATACTACC TATTCGCTCA 2760

CAATAAGAGA GAACTTTTTT ACGAAAATCT ATTGAATATG CCATAAGAAG ATTATACCAC 2820

ATTGTGTACT ATTTTTGGTT CATTTCACTA TAACACAAAA TAGATTATTA TTACATAACA 2880

AAAAAGAGGT CTAAACCTCT TAACTCAATT ACTCCGCCAG TAGGACTCGA ACCTACGACA 2940

TCATGATTAA CAGTCATGCG CTACTACCAA CTGAGCTATG GCGGATTAAA GCTAAGCGAC 3000

TTCCCTATCT CACAGGGGGC AACCCCCAAC TACTTCCGGC GTTCTAGGGC TTAACTTCTG 3060

TGTTCGGCAT GGGTACAGGT GTATCTCCTA GGCTATCGTC ACTTAACTCT GAGTAATACC 3120

TACTCAAAAT TGAATATCTA TTCAATTTAA GAAAACCGTT CGCTTTCATA TTCTCAGTTA 3180

CTTTGGATAA GTCCTCGAGC TATTAGTATT AGTCCGCTAC ATGTGTCGCC ACACTTCCAC 3240

TTCTAACCTA TCTACCTGAT CATCTCTCAG GGCTCTTACT GATATATAAT CATGGGAAAT 3300

CTCATCTTGA GGTGGkT:CA CACTTAGATG CTTTCAGCGT TTATCCCTTC CCTACATAGC 3360

TACCCAGCGA TGCCTTTGGC AAGACAACTG GTACACCAGC GGTAAGTCCA CTCTGGTCCT 3420

CTCGTACTAG GAGCAGATCC TCTCAAATTT CCTACGCCCG CGACGGATAG GGACCGAACT 3480

GTCTCACGAC GTTCTGAACC CAGCTCGCGT 3510

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20986 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

CGGAGAAAAA CATGGCTAAG TCAAACTTTG AAAAAGTAGA ATCAGTTGTT GGCTGGGTTC 60

GTGATAAGAA AATCACAGGC TACCGTATCT CTAAAGAAAC GAATGCGCGT GAAATGTCTA 120

TCATTGCTCT GGCGCAGGGT CGTGCAAAAG TAAAAAATAT TTCATTTGAA ACAGCCCTAG 180

GCCTAATTGA TTTCTATGAA AAAAATTATG AAAAATTTGA AGATTAATCT TTGGATAACG 240

GCGGATTCTT GACCTTCAAG TAGTAGAGAT AGAGAATCTG CCTTTTCATT TTGAGGACAG 300

CAAAAAGACT GCACGGTTGA TGCAGCCTTT TCTTTTTATT TGAGATAGCG TTGAAGGAAC 360

TCTTTTGTTC GGTCTTCTTT AGGATTGGTG AAGAGGTCTT CTGGTTTACC TTCTTCAGCG 420

ATCACGCCCT TATCCATAAA GATAACACGG TGAGAGACAT CACGGGCAAA TTCCATTTCA 480

TGGGTTACGA CAATCATGGT CAAGCCTTCC TGAGCCAGGT CCTGCATGAT TTTGAGGACT 540

TCTCCAACCA TTTCTGGATC GAGAGCTGAT GTTGGTTCAT CAAAGAGAAT AGCGTCCGGA 600

TTCATGGAGA GGGCACGAGC GATGGCCACA CGTTGTTTTT GACCACCTGA GAGTTGTTTT 660

GGTTTGGCTT GCCAGTAGCG TTCTCCCATG CCGACCTTTT CCAGGTTTTC TTTGGCAATC 720

TTTTCAGCTT CTGTGCGTTC GCGTTTTAGG ACAGTTGTCT GAGCGACGAT TGTGTTTTCA 780

AGAACATTGA GATTTTCAAA GAGGTTAAAG GATTGGAAAA CCATCCCCAA CTTTTCACGG 840

TATTGCGTGA GGTCATAGCC TTTTTCGAGG ACGTTTTGTC CATGATAAAG GATTTGTCCA 900

TCAGTTGGTG TTTCAAGTAG GTTAATGGAG CGTAGGAAGG TCGATTTTCC GCTTCCAGAG 960

CTTCCGATGA TAGAGATGAC CTCTCCCTTG TGGACAGTGA GTGAAATGTC TTTTAGCACT 1020

TCGTTTTGTC CATAGGATTT TTTGAGGTGT TTAATTTCAA GGATTGCTTG TGTCATTATT 1080

TCAAATCCTC CGTTTGCATT TGGTTAGCAC CTGTAGTGTA GGTATCCATG TCCATTCTGC 1140

GCTCGATAAA GCGTAGGATA CGTGTTACGG TGAAGGTGAG GACAAAGTAA ATCACGGCGA 1200

TGATTGTAAA TGTCTGGAAG TATTGATAGG TTTGTGTTGC CACGGTATTT CCTGAGAAAT 1260

AAAGTTCGAC AACAGAGATA ACGTTCAATA CAGATGTATC TTTGATATTG ATGACAAATT 1320

CATTACCAGT TGCAGGTAGG ATGTTACGGA CTACCTGAGG TAGGACAATC TTACGCATGG 1380

TCTGGTTATG GGTCATACCA AGAGCAGTCG CAGCTTCAAA TTGTCCCTTG TCAACTGCTA 1440

GGATACCACC ACGGACGATT TCAGTCATGT AGGCACCGGT ATTGATTGAA ACGATGAAGA 1500

TAGCAGCCAG TGTACGGTCA AGGTTGATCC CGAAAGCTTG GGCAGTTCCA TAGTAGATAA 1560

CCATCGATTG AACAATCATT GGCGTACCAC GGAAAATTTC AATGTAGACA TTGAGAACCC 1620

AGCCGACTAG TTTTTGTAGG CCGTAAATGA CTTTGTTTTC AGAGAGAGGA GCAGTACGGA 1680

AGACACCAAT GGCAAGTCCA ATAATGAGAC CTATGATGGT TCCGACGATA GAGATTAAAA 1740

GAGTGATACC AGCACCACGC AAGAGTTGTT GCCAGTTTTC AGAAAGAATT TTAGCAACTT 1800

GGCTAAAGAA ACTACTGCTA GTCTCTTCAG TTGTTGTAGC TTCGGCAGGT TGTTCCTTGA 1860

TCATACGATC CATCAAGGCA ACTTGGTCAT CTTTTGAAAT GGTTTCAATG CTGGCATTGA 1920

TTTGGCTAAT ACGATTGTCA TTTTTACGAA GCCCGATAGC GATAGCTGTA TCTTCTTCCC 1980

CAGTTTTGAA ACCAGGTTCT ACTTGAATCA TCTTGAACTT AGAGTTCGCA GCTTCAGCAG 2040

TCAGTGCTTC TGGACGTTCA GAAACATAAG CATCAATGAC ACCAGCCTCA AGAGCTTGTC 2100

GCATTTGAGC GAAGTCTCCC ATGGCTGTTT CTTTTTTAGC ACCTGGGATT TGTGCAATCA 2160

AGTTATAAAG GTAGACCCCT TGTTGAGAAG TGATTTTTGC ACCGTTAAAG TCATCCAAAG 2220

ATTTAGCACT TGCGTAGGCA GAATCTTTTT TGACAAGCAA AACTGGTTCG CTAGTATAGT 2280

AACTGCTCGA AAAGGCAATT TCTTGTTTGC GTTCTGCAGT TGGACTCATA CCTGCGATAA 2340

TCATGTCAAT CTTACCAGAA GTAAGGGCAG GGACTAGACC TTCCCACTTG GTTTTAACAA 2400

CCAAAGGTTC TTTACCTAAG TCCTTAGCGA TTTTCTTGGC GATTTGAACA TCGTATCCGT 2460

TGGCATACTG ATTGGTCCCA TCGATTTTGA CAGCTCCGTT GCTATCATCA TCCTGGGTCC 2520

AGTTAAAGGG AGCATATGCT GCTTCCATAC CGATGCGTAA ATATTCATCG GCTTGAGCAA 2580

CATTGACAAG TCCTAGCATC AGCAAGAGAC TTGTGAAAAT AGATAAGTAy ATGTGGCTCA 2640

TGATTTCTCC TATTCTGATC TATTAAAAAA TAACTGTCTC CTATTTTATC GAAAAATGCG 2700

TAATTTTTCA ACATAAGTAA GTCTTTACTT ACGAAAAAAT GCTATAATGA TAAGAAAGAT 2760

AAAAAGGGGG CTTAGTTGAT GAAAAAAACT TTTTTCTTAC TGGTGTTAGG CTTGTTTTGC 2820

CTTCTTCCAC TCTCTGTTTT TGCCATTGAT TTCAAGATAA ACTCTTATCA AGGGGATTTG 2880

480

TATATTCATG CAGACAATAC GGCAGAGTTT AGACAGAAGA TAGTTTACCA GTTTGAGGAG 2940

GACTTTAAGG GCCAAATCGT GGGACTTGGA CGTGCTGGTA AGATGCCTAG CGGGTTTGAC 3000

ATTGACCCTC ATCCAAAGAT TCAGGCCGCG AAAAACGGTG CAGAACTAGC AGATGTGACT 3060

AGCGAAGTAA CAGAAGAAGC GGATGGTTAT ACTGTGAGAG TCTATAATCC AGGTCAGGAG 3120

GGCGACATAG TTGAAGTTGA CCTCGTCTGG AACTTAAAAA ATTTACTTTT CCTTTATGAT 3180

GATATCGCTG AATTAAATTG GCAACCTCTG ACAGATAGTT CAGAGTCTAT TGAAAAGTTT 3240

GAATTTCATG TAAGGGGAGA CAAGGGGGCT GAAAAACTCT TTTTCCATAC AGGGAAACTT 3300

TTTAGAGAGG GAACGATTGA AAAGAGTAAC CTTGATTATA CTATCCGTTT AGACAATCTT 3360

CCGGCTAAGC GTGGAGTTGA GTTGCATGCC TATTGGCCTC GGACCGATTT TGCTAGCGCT 3420

AGGGATCAGG GATTGAAAGG GAATCGTTTA GAAGAGTTTA ATAAGATAGA AGACTCGATT 3480

GTTAGAGAAA AAGATCAGAG TAAACAACTC GTTACTTGGG TCCTCCCTTC GATCCTTTCC 3540

ATCTCCTTGT TATTGAGTGT CTGCTTCTAT TTTATTTATA GAAGAAAGAC CACTCCTTCA 3600

GTCAAATATG CCAAAAATCA TCGTCTCTAT GAACCACCAA TGGAATTAGA GCCTATGGTT 3660

TTATCAGAAG CAGTCTACTC GACCTCCTTG GAGGAAGTGA GTCCCTTGGT CAAGGGAGCT 3720

GGAAAATTCA CCTTTGATCA ACTTATTCAA GCTACCTTGC TAGATGTGAT AGACCGTGGG 3780

AATGTCTCTA TCATTTCAGA AGGAGATGCA GTTGGTTTGA GGCTAGTAAA AGAAGATGGT 3840

TTGTCAAGCT TTGAGAAAGA CTGCCTAAAT CTAGCTTTTT CAGGTAAAAA AGAAGAAACT 3900

CTTTCCAATT TGTTTGCGGA TTACAAGGTA TCTGATAGTC TTTATCGTAG AGCCAAAGTT 3960

TCTGATGAAA AACGGATTCA AGCAAGAGGG CTTCAACTCA AATCTTCTTT TGAAGAGGTA 4020

TTGAACCAGA TGCAAGAAGG AGTGAGAAAA CGAGTTTCCT TCTGGGGGCT CCCAGATTAT 4080

TATCGTCCTT TAACTGGTGG GGAAAAGGCC TTGCAAGTGG GTATGGGTGC CTTGACTATC 4140

CTGCCCCTAT TTATCGGATT TGGTTTGTTC TTGTACAGTT TAGACGTTCA TGGCTATCTT 4200

TACCTCCCTT TGCCAATACT TGGTTTTCTA GGGTTAGTTT TGTCTGTTTT CTATTATTGG 4260

AAGCTTCGAC TAGATAATCG TGATGGTGTT CTAAATGAAG CGGGAGCTGA GGTCTACTAT 4320

CTCTGGACCA GTTTTGAAAA TATGTTGCGT GAGATTGCAC GATTGGATCA GGCTGAACTG 4380

GAAAGTATTG TGGTCTGGAA TCGCCTCTTG GTCTATGCGA CCTTATTTGG CTATGCGGAC 4440

AAGGTTAGTC ATTTGATGAA GGTTCATCAG ATTCAAGTGG AAAATCCAGA TATCAATCTC 4500

TATGTAGCTT ATGGCTGGCA CAGTACGTTT TATCATTCAA CAGCACAAAT GAGCCATTAT 4560

GCTAGTGTCG CAAATACAGC AAGCACCTAC TCTGTATCTT CTGGAAGTGG AAGTTCTGGT 4620

GGTGGCTTCT CTGGAGGCGG AGGTGGCGGC AGTATCGGTG CCTTTTAAAG AGAGCTACCA 4680

TAGACTGAAA AAGTATGATA TAATGGAAGA TAGAAAAAAG ACAAACTATA AGAAAAGTCA 4740

ATAGTTTTAT CTAAACTATT TCTTATTTCA ATTTGATGAT TTGGCGATGA TTTTAGAGCA 4800

CGGCAAAAAG CCCTTGAAAA AGTCCATTTT TTCAAAGGTA ATCCTGTGTT AATTTCAGAA 4860

ATTACATCAC TTTTTGTTCG TCAAATGGCA GCTCTTTTTT AGGATATAAA ACAGGGTTCG 4920

GATAAGTTTT TTTGCAAGGT GGATGATGGC TACATTGTAA TGTTTTCCTT ATTCTAACTT 4980

AGTCTTAAGA TAGGCCTTAG AAGCAGGTGA AAAGCGAGGG CATGCTTTGG CAGCTTGTAT 5040

GAGTGCCCAC CGCAGATGAG GGGAACCCCG TTTGACCATT CTTCCAGCTA AATCAATCTG 5100

ACCTGACTGA TAAATAGAAG AATCCAGTCC AGCGAAAGCT TGTAATTGAG CAGGATTATC 5160

AAAGGCATGA ATATTTCGAA TCTCGGCTAA AATGACCGCC CTAAACGATC CCCAATCCCA 5220

GTAACCGTCG TGATGACCGA GTTGAACTCA GCCATCGAGT CATTGATACA TGTTTCCGCC 5280

TTGTCAATGA GCCTCTTGTA ATGCTTGATG ATTTCGAATT CACGAGCAGG AGATGTTGTT 5340

CCGATAGAAC GAGGTGCGAC TGAGAGGATA TCCTGAATTT TAGAAGCGGT CAATCGCTTA 5400

ATTTCTATCA GCTTATCAAA TCCTGCCTCA ATCCTTTTCT GAGGATTAGG GTAGCGTGTC 5460

AAGAGTTGGT AGGTATATTC TGAATGCTTT CCAACGATTT TATCCAACTC AGGAAAGATG 5520

ATATCAAGAC AACGAGTGTA TTGTACTTTC CAATCAGACT GTTTTTCTTG AGACGATGAA 5580

TATGTCTAGC CAGTATTTTT AGGTCTACTT GCCGATTATC GTGTTGAAAT TGTTCACGAT 5640

TGGGGTCAGA AAGAAGTTTA AGAGCGATGC CATGAGCGTC TTTCTTATCC GTTTTAGTCT 5700

TGCGAAGTCA TAATGATTTG GCAAATTCCT TGATGAGCAA AGGATTGTAG GTGTAAACTT 5760

TATATCCTTG TTCATGCAGG AAGTTCAGTA GATTAAAGGC ATAATGTCCA GTATCTTCAA 5820

GAGCGATGAG ACAGTCTTGG TTGATCTGTC GAATAGACAG ATCTAAGAGT TCAAAACCAG 5880

CTTTATTATT TGAAAAAGTG AGTGGTTTAA GAACAGTTTT TCCTGGAACA TTCAAGGCTG 5940

TAACATCGTG TTTATTTTTA GCGATATCAA TGCCTACATA AAGCATGGGA GTACCTCCAG 6000

ATATAGTATT TCAAGTCTAC TTGGTTATCC ACGAATTTTT TGCCTTGTTA CCTTAGACGA 6060

GATCAAACGT CTATGCGTTA TCAAACTCAT TACCAATTGA AACAAAAGCT GTGGTTAGAG 6120

CCTTTCGGAA ATCGTCAAGC GATTGGAGGA AATGAACTAA TCCATAGTGG CTTATTCCAA 6180

GTATACCACT TGGGCTTTGG CAGTAGCTAA CTGCGCTAAA TATAATATAG GGAGTAATCT 6240

ATGTATCTTA TTGAAATTTT AAAATCTATC TTCTTCGGAA TTGTTGAAGG AATTACGGAA 6300

TGGTTGCCGA TTTCCAGTAC AGGTCACTTG ATTTTAGCAG AGGAATTCAT CCAATACCAA 6360

AATCAAAATG AAGCCTTTAT GTCCATGTTT AATGTCGTGA TTCAGCTTGG TGCTATTTTA 6420

GCAGTTATGG TGATTTATTT TAACAAGCTC AATCCTTTTA AACCGACCAA GGACAAACAG 6480

GAAGTTCGTA AGACTTGGAG ACTATGGTTG AAGGTCTTGA TTGCTACTTT ACCTTTACTT 6540

GGTGTCTTTA AATTTGATGA TTGGTTTGAT ACCCACTTCC ATAACATGGT TTCAGTTGCT 6600

CTCATGTTGA TTATCTACGG GGTTGCCTTC ATCTATTTGG AAAAGCGCAA TAAAGCGCGT 6660

GCTATCGAGC CAAGTGTAAC AGAGTTGGAC AAGCTTCCTT ATACGACCGC TTTCTATATC 6720

GGACTCTTCC AAGTTCTTGC TCTTTTACCA GGGACTAGCC GTTCAGGTGC AACGATTGTC 6780

GGTGGTTTGT TAAATGGAAC CAGTCGTTCA GTTGTGACAG AATTTACCTT CTATCTTGGG 6840

ATTCCTGTTA TGTTTGGAGC TAGTGCCTTA AAGATTTTCA AATTTGTGAA AGCCGGAGAA 6900

CTCTTGAGCT TTGGGCAATT GTTTTTGCTC TTGGTCGCGA TGGGAGTAGC TTTTGCGGTC 6960

AGCATGGTGG CTATTCGCTT CTTGACCAGC TATGTGAAAA AACACGACTT CACCCTTTTT 7020

GGTAAATACC GTATCGTGCT TGGTAGTGTT TTGCTACTTT ACAGTTTTGT CCGTTTATTT 7080

GTATAAGAAA AACCTTGAAG GGGCAACTCT TCAAGGTTTT ATACTCTTCG AAAATCTCTT 7140

CAAACCGCGT CAGCTTTATC TGCAACCTCA AAACAGTGTT TTGAGCAGCn CTGCGGCTAG 7200

CCTCCTAGTT TGCTCTTTGA TTTTCATTGA GCTTTAAAAT CCAGTCATGG TAATCCCCAA 7260

TAGGCGGACA CCTCTTTCTT TCTTGCTTAA TTCTTCATAG AGTTGCAGGG CTATTTGGCT 7320

TATCTGACTA GCATCTTGTG TTTTTTGAGC AAGACTTTTT CGTTTGGTAA GAGTTGAAAA 7380

GTCCTCGTAG CGGATTTTCA AAATGACAAT TTTTCCAGCT TTTTCTTGTT GATGTAGATT 7440

GAGAGCGACT TTTTCTGATA GAAGAGTCAG CTCTTTTTTG ATATCTTCCT CAGCAAGGAG 7500

AATCTTCCCG TAGGTTTTCT CCTTGCCGAT TGATTTACGG ATGCGATTGG ATTTGACTGG 7560

AGAGTTGTGA ATGCCACGAG CCTTTCGATA CAGATCATAG CCTAGTCTAC CAAAACGGTC 7620

TATTAGGGTT ACCTCAGGAA CTTCAAGTAA ATCAGCACCA GTAAAAACGC CCATTTGATG 7680

AAGACGTTCT ACTGTCTTTT TTCCTACTCC ATGAAATTTG GAAATATCCA TTTGTTTGAG 7740

AAAATCCTCA GCCTGTTCAG GTAGAATCAC TGTCAAACCA TGTGGTTTTT GATAATCACT 7800

CGCCATTTTA GCTAAGAATT TGTTGTAAGA AACGCCTGCG GAAGCAGTTA GATGGAGTTC 7860

TTGCCAGATA TCTTTTTGAA TGAGGCGAGC AATTTTGACC GCTGACTTGA TACCGAGTTT 7920

ATTTTCTGTC ACATCCAAAT AGGCTTCGTC AATGCTCATG GGTTCAATCA AATCTGTATA 7980

GCGCTTAAAA ATAGCTCGAA TCTGGAGTCC CACAGACTTG TATTTCTCAT AATTCCCTGA 8040

GATAAAGACA GCCTGGGGAC AACGTTCATA AGCTTCCTTG GAACTCATGG CAGAATGGAC 8100

ACCAAAAGCT CTTGCCTCAT AACTACAGGT AGAAACGACT CCCCGTCCAC CTGTTTGCCG 8160

AGGGTCGCTT CCAATAATGA CAGGTTTTCC TCTGAGTTTA GGATTATCCC TGATTTCCAC 8220

TGCAGCAAAA AAGGCATCCA TGTCAATATG GATGATTTTT CTTGACAAAT CATTTAACAA 8280

AGGAAAAATC AACATGCCTA GCACCTTTTT ATACTCTTCG AAAATCTCTT CAAACCACGT 8340

CAGCTCTATm TGCAACCTCA AAACAGTGTT TTGAGCAATC TGCGGCTAGC TTCCTAGTTT 8400

GCTTTTCGAT TTCCATTGAG TGTTACTGCT TATTyTCTTT TATTATACCC TTTTTTCTGA 8460

AAAAAAGAAA AAAGGACTTT ATTTTTTCAA AAATATAATA CAGTTTGAAA TAAAATATAG 8520

ACTGTTTTAG AAAAGAAAGT GTAAAAATAG GGAATTTTCA CTTGTTGAAA TCGGTTACTA 8580

TATGGTATAC TTGTCTTATG AATGTAACAG ATGACTGTTA CTAGAAAAAA GAGGACATTA 8640

ATATGGTTGT TAAGACAGTT GTTGAAGCAC AAGATATTTT TGACAAAGCT TGGGAAGGCT 8700

TCAAAGGCGT AGATTGGAAA GAAAAAGCAA GTGTATCACG ATTTGTACAA GCTAACTACA 8760

CACCTTATGA TGGAGACGAA AGCTTCCTTG CAGGACCAAC AGAGCGTTCA CTTCACATCA 8820

AGAAAATTGT AGAAGAAACT AAAGCACACT ACGAAGAAAC TCGTTTCCCA ATGGACACTC 8880

GTCCAACATC TATCGCTGAT ATCCCTGCTG GATTTATCGA CAAAGAAAAT GAAGTTATCT 8940

TCGGTATCCA AAACGATGAA CTCTTCAAAT TGAACTTCAT GCCAAAAGGT GGTATCCGTA 9000

TGGCTGAAAC TACTTTGAAA GAAAATGGAT ACGAACCAGA CCCAGCTGTT CACGAAATCT 9060

TCACTAAATA TGTAACAACA GTTAACGACG GTATTTTCCG TGCCTACACT TCAAATATTC 9120

GTCGCGCTCG TCACGCACAC ACTGTAACTG GTCTTCCAGA TGCATACTCA CGCGGACGTA 9180

TCATCGGTGT TTACGCACGT CTTGCTCTTT ACGGTGCAGA CTACTTGATG CAAGAAAAAG 9240

TAAATGACTG GAATGCAATC AAAGAAATCG ATGAAGAAAC AATCCGTCTT CGTGAAGAAG 9300

TAAACCTTCA ATACCAAGCA TTGCAACAAG TTGTTCGCCT GGGTGACCTT TACGGGGTTG 9360

ATGTTCGCAA ACCAGCGATG AACGTGAAAG AAGCAATCCA ATGGGTTAAC ATTGCTTTCA 9420

TGGCTGTCTG CCGTGTGATT AACGGTGCTG CTACATCTCT AGGTCGTGTA CCAATCGTAT 9480

TGGACATCTT TGCAGAACGT GACCTTGCTC GTGGTACATT TACTGAATCA GAAATCCAAG 9540

AATTCGTTGA TGATTTCGTT ATGAAACTTC GTACAGTTAA ATTTGCTCGT ACAAAAGCTT 9600

ATGACCAATT GTACTCAGGT GACCCAACCT TTATCACAAC TTCTATGGCT GGTATGGGTA 9660

ACGACGGTCG TCACCGTGTT ACTAAGATGG ACTACCGTTT CTTGAACACT CTTGACAACA 9720

TCGGTAACTC ACCAGAACCA AACTTGACAG TTCTTTGGAC TGACAAATTG CCATACAACT 9780

TCCGTCGCTA CTGTATGCAC ATGAGCCACA AACACTCTTC TATCCAATAC GAAGGTGTAA 9840

CAACAATGGC TAAAGACGGA TATGGTGAAA TGAGCTGTAT CTCATGCTGT GTGTCTCCAC 9900

TTGATCCAGA AAATGAAGAA CAACGCCACA ACATCCAGTA CTTCGGTGCT CGTGTAAACG 9960

TTCTTAAAGC CCTTCTTACT GGTTTGAATG GTGGTTACGA CGATGTTCAC AAAGACTACA 10020

AAGTATTTGA TATCGAACCA ATCCGTGACG AAGTTCTTGA ATTTGAATCA GTTAAAGCGA 10080

ACTTTGAAAA ATCTCTTGAC TGGTTGACTG ACACTTACGT AGATGCCTTG AACATCATCC 10140

ACTACATGAC TGATAGGTAC AACTACGAAG CTGTTCAAAT GGCCTTCTTG CCAACTAAAC 10200

AACGTGCCAA CATGGGATTC GGTATCTGTG GATTTGCTAA CACTGTTGAT ACATTGTCAG 10260

CTATCAAATA CGCTACAGTT AAACCAATCC GTGACGAAGA TGGCTACATC TACGATTACG 10320

AAACAATCGG TGACTACCCA CGCTGGGGTG AAGATGACCC ACGTTCAAAC GAATTGGCAG 10380

AATGGTTGAT CGAAGCTTAC ACAACTCGTC TACGTAGCCA CAAACTATAC AAAGACGCAG 10440

AAGCTACAGT ATCACTTTTG ACAATCACAT CTAACGTTGC TTACTCTAAA CAAACTGGTA 10500

ACTCACCAGT TCACAAAGGT GTATACCTCA ACGAAGATGG TTCTGTGAAC TTGTCTAAAC 10560

TTGAATTCTT CTCACCAGGT GCTAACCCAT CTAACAAAGC TAAAGGTGGT TGGTTGCAAA 10620

ACTTGAACTC ACTTTCTAGC CTTGACTTTA GTTATGCAGC TGACGGTATC TCATTGACTA 10680

CACAAGTATC ACCTCGCGCT CTTGGTAAGA CTCGTGATGA ACAAGTTGAT AACTTGGTAA 10740

CAATTCTTGA TGGTTACTTC GAAAACGGTG GACAACACGT TAACTTGAAC GTTATGGACT 10800

TGAACGATGT TTACGAAAAA ATCATGTCAG GCGAAGACGT TATCGTACGT ATCTCTGGAT 10860

ACTGTGTAAA CACTAAATAC CTCACTCCAG AACAAAAAAC TGAATTGACA CAACGTGTCT 10920

TCCACGAAGT TCTTTCAATG GATGACGCCT TGGATGCATT GAGCTAATCA AGTTCTTGAA 10980

TAATAAAAAG GAACCCTCGG TCAAACGACT GAGGGTTTTG TGCTTGGGAT AGTATGAGCA 11040

ATTCCTTCGG CGCAATATGC AATGTTTTTG GGCTCTTTGT CAACTGTAGT GGGTTGAAAA 11100

AAAGCTAAGC TTGAGAAAGG ACAAATTTCG TCCTTTCTTT TTTGATGTTC AGGGCGATAA 11160

AAATCCGTTT TTTGAAGTTT TCAAAGTTCC GAAAACCAAA GGCATTGCGC TTGATGTCTT 11220

TGATGAGTTT GTTAGTGGCC TCAAGTTTAG CGTTAGAATA AGGCAATTCA ATGGCGTTAG 11280

TGATGTAGTT TTTATAGCAA ATAAATGTGC TCAAAGTGGT TTTAAAGGTG CGGTTGAGAT 11340

GAGGTAACGT GTCTTGAATT AAGCCCCAAA ACTGGTCAGT ATTCTTCTCT TGTAGATGAA 11400

ATAGGAGTAG TTGATACAGG TCATAGTAAT CTTTAAGTTC AGGTACTAGA GTAAAGATTT 11460

TCTTCAGACA CTCCCTAGGA GTTAAGGTCT CTCTGAAAGT TCTAGCATAG AAAGGCTTAA 11520

GAGAGAGTTT CCGACTATCT TTTAGGATAA ATTTCCAGTA ATATTTAAGA GCTCTGTATT 11580

CCAGAGATTT ATCATCAAAT TGCTTCATGA TGTTGATTCT AGTCTGATTA AGAGCCCTGC 11640

TCATGTGTTG GACAATGTGG AAACGATCGA GAACAATTTT AGCATTGGGA AATAATTTCT 11700

TAATGAGAGG GATATAACTT CCAGACATAT CAACAGTGAC GACTTTAACT TTTTTTCTAG 11760

CTTCTTTCGA GTACTTGAAG AAATGATTTC GGATGGTTGT TTGACGTCTG TTATCAAGAA 11820

TGGTCATGAT TTTCTTAGTG TTGAAATCCT GAGCAATGAA AGCCAATTTC CCCTTCTGGT 11880

AGGAGAATTC ATCCCAGGAG AGGATTTCAG GCAAAGTGGT GTAATCCTCT TGGAAATGAA 11940

ATTGCTTGAG CTTACGATAG ACGGTAGAGG TAGAGGTAGA GGTAGAGATG GCTAATTTAG 12000

AAGCGATATG TGTAAGAGCC TCTCTGTTGA GTAGGAGTTG GGCAATTTTC TGTCTCACCA 12060

TTTCCGAGAT TTGGCAATTT TTCTGAACGA GAGTTGTTTC AGCTACAGTG ACTTTCCGAC 12120

AGGACTTGCA TTGAAATCGT CTCTTTTTCA AATGAATGAG GCTAGGGAAA CCACCAATCT 12180

CGATAAAAGG GATTTTAGAA GGCTTTTGGA AGTCGTATTT GATTTGTTTT CCTTTACAGT 12240

GTTTACATTT AGGTGGGTGA TAATCAAGTG TAGCGAAGAC TTCGATATGG GTATCGTGCT 12300

GAATGGCTTT ATTTAAGGTG ATGTTTTTGT CTTTTATTCC GATGAGTAAT GTGGTATGAT 12360

TGATGTGTTC CATAAGATAC TTTCTAATGA GTTGTTTAGG CGCTTTTCAT TATAAGTCTT 12420

ATGGGACTTT TTTGATACTC AAAAAGCCCT ATAATCTCCA CAGTGGGATT TACCCACTAC 12480

AGAAATTATA GAGCCAGAAA AAACACTTTT GTTCACTAGC AGAAACTAGA GAGCAGAAGT 12540

GTTTTTCTGT TCAGATTTAC CCAAAACTGG GAAATATGGG GATAAGAATA GAGATGGCTT 12600

AGGAAGCCCC TTTTTGTGTG TAGACAGTAC GATGAACTTA TAACAAATAG TGAGCCTTTT 12660

TAGCAATCAT TGCGACCCGT TTGTCAAAAG CCTCTTTTCG GATATCTACA ATTGTCTGAT 12720

AGATGAGACG CTGTTGGCTA ACATGCAAAT CTAAGGCAAT CGTCAAAAAG TGATGTTTCC 12780

CTTTGGGATA CTGCTTTTTA ACGTAAGGCA GGTATTCTTT CGTTGTAATA ATAATCAATG 12840

GCTCTGTCAA ATGCTCCTCT GAAGGAGGAG GACTAATTAG AATATTGTAT CCTGTAACAG 12900

AGGCAACTTT GTCAGTAAAA TTCCGTAAAA TAATGGACTT TATTAAGTTT ACATCTGCTT 12960

GATTATTTAA AATGATAAAA ATCGGGATAG CAGGTAGTGA GGAAAAGATG GTTTCTGTCA 13020

AGTAGAGTGA GAAAAGGTAC AGCCGATGCT GGTCGATAAC TCCTTCAATC TTCTGCTCAG 13080

TCATCCACTC TTGAACAATT GCTTTCGAAA TATGATACAG TGGCTTGTCG CTTTCAATCC 13140

CATAATGTTC GTAATAATTA TAATAGGGAA CTAGATTTTG TAAACCAAAC AAAAACGTTC 13200

TTGTTAAGAA AGTCAGTGCT GTTAAAAAAG AAAGAGAATT CGAAATGTCA TTTCCTAAGA 13260

TATTCTTGAA CTTGGATAGT AGATGCTTTC CTCTTGTATG CTGAAGAATC AGTTGAATAG 13320

TATGAGTCTT TTTTTCTTGA TTCCATTTGT CCTTGGAAAA CGAAGAATTA GCAGAACAAT 13380

AAACCAAAAA GATATAATCC AGTTCTTCCT GAGTAAAAGT CATGTTGGCA TGTGGCTCTA 13440

AGTAAGTTTG GCAATGTTCC ATCAAAATCG GATACATAAA GAGGTTTTTT AATTTTTCAA 13500

ACTCTTTGGA CTCAGGGAAC TCAAGTGGAA ATTCCCGACG TTTCCAAGTG AGTGCCACTA 13560

GTATGCTAAA ATGAACATAC TCGTCAGGTG TGATTTCTAA CAGTTCATGA CTGAGTTGAG 13620

AATTAGACTG CACAATCATA TGTGTGACCC AATCCATACT TCCATCATTC AAATCATAAA 13680

TCTCAATACC AAAATGAAAC TGGAGGAGTG CAATTAAAAA ACGAATGCGA TATTCAGGAC 13740

CAACTACTTG ATTTTTCACA AGGTCCAAAC CTACTGAACG TAGTAACAAG CCACACTTTT 13800

GTCGTACGCG GTAGCCTGTT GCGATGGAAA TATACTCTTT TTGTGTAAAT TCGTTAAAGC 13860

TTTGATTACC TTGTAGTAGA AAGAAGCGGA GTATTTTTAA AATAGTTGAT TGGTTATAAA 13920

GCTGATGGAA GTAATAATTC GTTTGATGAG AATGGTGTTC GATTAATTGA ACTTGTTGCG 13980

TATCTAAATT AAATGTCAAC TCTTCCTCGA ATGTTTCTTG TAATTCCTGC AAAATGCTTA 14040

GGAGACTTTT AGATTGTAAT GAAGTTAAAG TAGACAGTTC ATCTAGTTCA ATAGACCGAA 14100

TATCCAATAA TATATTTAAA ATGGTAATTT TATCTGTAAT TCTTTTTTCA ATGTATTTGT 14160

TTAGCATAGT TACCGAATCT TAGTTGCATA TAGATAATTT TAATTATTAT AATACAAAAG 14220

AAACTAATTG TCTTGTCAAA AAGGTTGTGG AATTTCCGAC TTTATTGATA AAACAGCATG 14280

TAATAAAAGG CATTTTAAAG ATAGTAATGA GTATTGGTGG AGTTTTATGG CTTATTTTTT 14340

TTATTAGAAA ATATTTTTTT ATCAAATATT GTCGTTCTAT AAAAAAAATAT GTGATAAAAA 14400

TATCTATTGT GATGGAAGTT GTTTTAATTT ATACTAGGAT AGTTAATAGT AATACTATAC 14460

TATACTATAT TGTATACAAG TGTGTCATTG CCAGGTTGAG AAGATAGCTA TAACGCACTT 14520

TTATACGCTT TTGCTACGTT TGTTAGTGAA CGGATTAACT CAGTGAGATA AATTTTATCA 14580

GAACATAAGT AATCCGTTTC TTCGTGTATA CAGATTGAAA GTACCTATGA ATCATAGAAG 14640

GATTAACTTG TTCTATGAAT AATGCTTAAC AGGGAGACAC ACATGAAAAA AGTAAGAAAG 14700

ATATTTCAGA AGGCAGTTGC AGGACTGTGC TGTATATCTC AGTTGACAGC TTTTTCTTCG 14760

ATAGTTGCTT TAGCAGAAAC GCCTGAAACC AGTCCAGCGA TAGGAAAAGT AGTGATTAAG 14820

GAGACAGGCG AAGGAGGAGC GCTTCTAGGA GATGCCGTCT TTGAGTTGAA AAACAATACG 14880

GATGGCACAA CTGTTTCGCA AAGGACAGAG GCGCAAACAG GAGAAGCGAT ATTTTCAAAC 14940

ATAAAACCTG GGACATACAC CTTGACAGAA GCCCAACCTC CAGTTGGTTA TAAACCCTCT 15000

ACTAAACAAT GGACTGTTGA AGTTGAGAAG AATGGTCGGA CGACTGTCCA AGGTGAACAG 15060

GTAGAAAAATC GAGAAGAGGC TCTATCTGAC CAGTATCCAC AAACAGGGAC TTATCCAGAT 15120

GTTCAAACAC CTTATCAGAT TATTAAGGTA GATGGTTCGG AAAAAAACGG ACAGCACAAG 15180

GCGTTGAATC CGAATCCATA TGAACGTGTG ATTCCAGAAG GTACACTTTC AAAGAGAATT 15240

TATCAAGTGA ATAATTTGGA TGATAACCAA TATGGAATCG AATTGACGGT TAGTGGGAAA 15300

ACAGTGTATG AACAAAAAGA TAAGTCTGTG CCGCTGGATG TCGTTATCTT GCTCGATAAC 15360

TCAAATAGTA TGAGTAACAT TCGAAACAAG AATGCTCGAC GTGCGGAAAG AGCTGGTGAG 15420

GCGACACGTT CTCTTATTGA TAAAATTACA TCTGATTCAG AAAATAGGGT AGCGCTTGTG 15480

ACTTATGCTT CCACTATCTT TGATGGGACC GAGTTTACAG TAGAAAAAGG GGTAGCAGAT 15540

AAAAACGGAA AGCGATTGAA TGATTCTCTT TTTTGGAATT ATGATCAGAC GAGTTTTACA 15600

ACCAATACCA AAGATTATAG TTATTTAAAG CTGACTAATG ATAAGAATGA CATTGTAGAA 15660

TTAAAAAATA AGGTACCTAC CGAGGCAGAA GACCATGATG GAAATAGATT GATGTACCAA 15720

TTCGGTGCCA CTTTTACTCA GAAAGCTTTG ATGAAGGCAG ATGAGATTTT GACACAACAA 15780

GCGAGACAAA ATAGTCAAAA AGTCATTTTC CATATTACGG ATGGTGTCCC AACTATGTCG 15840

TATCCGATTA ATTTTAATCA TGCTACGTTT GCTCCATCAT ATCAAAATCA ACTAAATGCA 15900

TTTTTTAGTA AATCTCCTAA TAAAGATGGA ATACTATTAA GTGATTTTAT TACGCAAGCA 15960

ACTAGTGGAG AACATACAAT TGTACGCGGA GATGGGCAAA GTTACCAGAT GTTTACAGAT 16020

AAGACAGTTT ATGAAAAAGG TGCTCCTGCA GCTTTCCCAG TTAAACCTGA AAAATATTCT 16080

GAAATGAAGG CGGCTGGTTA TGCAGTTATA GGCGATCCAA TTAATGGTGG ATATATTTGG 16140

CTTAATTGGA GAGAGAGTAT TCTGGCTTAT CCGTTTAATT CTAATACTGC TAAAATTACC 16200

AATCATGGTG ACCCTACAAG ATGGTACTAT AACGGGAATA TTGCTCCTGA TGGGTATGAT 16260

GTCTTTACGG TAGGTATTGG TATTAACGGA GATCCTGGTA CGGATGAAGC AACGGCTACT 16320

AGTTTTATGC AAAGTATTTC TAGTAAACCT GAAAACTATA CCAATGTTAC TGACACGACA 16380

AAAATATTGG AACAGTTGAA TCGTTATTTC CACACCATCG TAACTGAAAA GAAATCAATT 16440

GAGAATGGTA CGATTACAGA TCCGATGGGT GAGTTAATTG ATTTGCAATT GGGCACAGAT 16500

GGAAGATTTG ATCCAGCAGA TTACACTTTA ACTGCAAACG ATGGTAGTCG CTTGGAGAAT 16560

GGACAAGCTG TAGGTGGTCC ACAAAATGAT GGTGGTTTGT TAAAAAATGC AAAAGTGCTC 16620

TATGATACGA CTGAGAAAAG GATTCGTGTA ACAGGTCTGT ACCTTGGAAC GGATGAAAAA 16680

GTTACGTTGA CCTACAATGT TCGTTTGAAT GATGAGTTTG TAAGCAATAA ATTTTATGAT 16740

ACCAATGGTC GAACAACCTT ACATCCTAAG GAAGTAGAAC AGAACACAGT GCGCGACTTC 16800

CCGATTCCTA AGATTCGTGA TGTGCGGAAG TATCCAGAAA TCACAATTTC AAAAGAGAAA 16860

AAACTTGGTG ACATTGAGTT TATTAAGGTC AATAAAAATG ATAAAAAACC ACTGAGAGGT 16920

GCGGTCTTTA GTCTTCAAAA ACAACATCCG GATTATCCAG ATATTTATGG AGCTATTGAT 16980

CAAAATGGCA CTTATCAAAA TGTGAGAACA GGTGAAGATG GTAAGTTGAC CTTTAAAAAT 17040

CTGTCAGATG GGAAATATCG ATTATTTGAA AATTCTGAAC CAGCTGGTTA TAAACCCGTT 17100

CAAAATAAGC CTATCGTTGC CTTCCAAATA GTAAATGGAG AAGTCAGAGA TGTGACTTCA 17160

ATCGTTCCAC AAGATATACC AGCGGGTTAC GAGTTTACGA ATGATAAGCA CTATATTACC 17220

AATGAACCTA TTCCTCCAAA GAGAGAATAT CCTCGAACTG GTGGTATCGG AATGTTGCCA 17280

TTCTATCTGA TAGGTTGCAT GATGATGGGA GGAGTTCTAT TATACACACG GAAACATCCG 17340

TAAAGTGTAG AAATGATAAT ATCTATGTTC TGAACGATAC TTTTAAGAAG TAGCACTCAA 17400

GAAGAGATTT AAGTTTACTT GGTGAAACCT GTTTTATTCG TAAGTAAACT ATCATTGAAA 17460

GGGGAGATGT TTTCGAAAAC TTGCACAGAA AAAGGATTAT TATTGTCATG TGTAATTCAT 17520

TACATTGCTC ACAGTTGATT TTAAGAGATA TGAATAAGGA GAAATCATGA AATCAATCAA 17580

CAAATTTTTA ACAATGCTTG CTGCCTTATT ACTGACAGCG AGTAGCCTGT TTTCAGCTGC 17640

AACAGTTTTT GCGGCTGGGA CGACAACAAC ATCTGTTACC GTTCATAAAC TATTGGCAAC 17700

AGATGGGGAT ATGGATAAAA TTGCAAATGA GTTAGAAACA GGTAACTATG CTGGTAATAA 17760

AGTGGGTGTT CTACCTGCAA ATGCAAAAGA AATTGCCGGT GTTATGTTCG TTTGGACAAA 17820

TACTAATAAT GAAATTATTG ATGAAAATGG CCAAACTCTA GGAGTGAATA TTGATCCACA 17880

AACATTTAAA CTCTCAGGGG CAATGCCGGC AACTGCAATG AAAAAATTAA CAGAAGCTGA 17940

AGGAGCTAAA TTTAACACGG CAAATTTACC AGCTGCTAAG TATAAAATTT ATGAAATTCA 18000

CAGTTTATCA ACTTATGTCG GTGAAGATGG AGCAACCTTA ACAGGTTCTA AAGCAGTTCC 18060

AATTGAAATT GAATTACCAT TGAACGATGT TGTGGATGCG CATGTGTATC CAAAAAATAC 18120

AGAAGCAAAG CCAAAAATTG ATAAAGATTT CAAAGGTAAA GCAAATCCAG ATACACCACG 18180

TGTAGATAAA GATACACCTG TGAACCACCA AGTTGGAGAT GTTGTAGAGT ACGAAATTGT 18240

TACAAAAATT CCAGCACTTG CTAATTATGC AACAGCAAAC TGGAGCGATA GAATGACTGA 18300

AGGTTTGGCA TTCAACAAAG GTACAGTGAA AGTAACTGTT GATGATGTTG CACTTGAAGC 18360

AGGTGATTAT GCTCTAACAG AAGTAGCAAC TGGTTTTGAT TTGAAATTAA CAGATGCTGG 18420

TTTAGCTAAA GTGAATGACC AAAACGCTGA AAAAACTGTG AAAATCACTT ATTCGGCAAC 18480

ATTGAATGAC AAAGCAATTG TAGAAGTACC AGAATCTAAT GATGTAACAT TTAACTATGG 18540

TAATAATCCA GATCACGGGA ATACTCCAAA GCCGAATAAG CCAAATGAAA ACGGCGATTT 18600

GACATTGACC AAGACATGGG TTGATGCTAC AGGTGCACCA ATTCCGGCTG GAGCTGAAGC 18660

AACGTTCGAT TTGGTTAATG CTCAGACTGG TAAAGTTGTA CAAACTGTAA CTTTGACAAC 18720

AGACAAAAAT ACAGTTACTG TTAACGGATT GGATAAAAAT ACAGAATATA AATTCGTTGA 18780

ACGTAGTATA AAAGGGTATT CAGCAGATTA TCAAGAAATC ACTACAGCTG GAGAAATTGC 18840

TGTCAAGAAC TGGAAAGACG AAAATCCAAA ACCACTTGAT CCAACAGAGC CAAAAGTTGT 18900

TACATATGGT AAAAAGTTTG TCAAAGTTAA TGATAAAGAT AATCGTTTAG CTGGGGCAGA 18960

ATTTGTAATT GCAAATGCTG ATAATGCTGG TCAATATTTA GCACGTAAAG CAGATAAAGT 19020

GAGTCAAGAA GAGAAGCAGT TGGTTGTTAC AACAAAGGAT GCTTTAGATA GAGCAGTTGC 19080

TGCTTATAAC GCTCTTACTG CACAACAACA AACTCAGCAA GAAAAAGAGA AAGTTGACAA 19140

AGCTCAAGCT GCTTATAATG CTGCTGTGAT TGCTGCCAAC AATGCATTTG AATGGGTGGC 19200

AGATAAGGAC AATGAAAATG TTGTGAAATT AGTTTCTGAT GCACAAGGTC GCTTTGAAAT 19260

TACAGGCCTT CTTGCAGGTA CATATTACTT AGAAGAAACA AAACAGCCTG CTGGTTATGC 19320

ATTACTAACT AGCCGTCAGA AATTTGAAGT CACTGCAACT TCTTATTCAG CGACTGGACA 19380

AGGCATTGAG TATACTGCTG GTTCAGGTAA AGATGACGCT ACAAAAGTAG TCAACAAAAA 19440

AATCACTATC CCACAAACGG GTGGTATTGG TACAATTATC TTTGCTGTAG CGGGGGCTGC 19500

GATTATGGGT ATTGCAGTGT ACGCATATGT TAAAAACAAC AAAGATGAGG ATCAACTTGC 19560

TTAAGTAAGA GAGAAAGGAG CCATTGATGA CAATGCAGAA AATGCAGAAA ATGATTAGTC 19620

GTATCTTCTT TGTTATGGCT CTGTGTTTTT CTCTTGTATG GGGTGCACAT GCAGTCCAAG 19680

CGCAAGAAGA TCACACGTTG GTCTTGCAAT TGGAGAACTA TCAGGAGGTG GTTAGTCAAT 19740

TGCCATCTCG TGATGGTCAT CGGTTGCAAG TATGGAAGTT GGATGATTCG TATTCCTATG 19800

ATGATCGGGT GCAAATTGTA AGAGACTTGC ATTCGTGGGA TGAGAATAAA CTTTCTTCTT 19860

TCAAAAAGAC TTCGTTTGAG ATGACCTTCC TTGAGAATCA GATTGAAGTA TCTCATATTC 19920

CAAATGGTCT TTACTATGTT CGCTCTATTA TCCAGACGGA TGCGGTTTCT TATCCAGCTG 19980

AATTTCTTTT TGAAATGACA GATCAAACGG TAGAGCCTTT GGTCATTGTA GCGAAAAAAA 20040

CAGATACAAT GACAACAAAG GTGAAGCTGA TAAAGGTGGA TCAAGACCAC AATCGCTTGG 20100

AGGGTGTCGG CTTTAAATTG GTATCAGTAG CAAGAGATGT TTCTGAAAAA GAGGTTCCCT 20160

TGATTGGAGA ATACCGTTAC AGTTCTTCTG GTCAAGTAGG GAGAACTCTC TATACTGATA 20220

AAAATGGAGA GATTTTTGTG ACAAATCTTC CTCTTGGGAA CTATCGTTTC AAGGAGGTGG 20280

AGCCACTGGC AGGCTATGCT GTTACGACGC TGGATACGGA TGTCCAGCTG GTAGATCATC 20340

AGCTGGTGAC GATTACGGTT GTCAATCAGA AATTACCACG TGGCAATGTT GACTTTATGA 20400

AGGTGGATGG TCGGACCAAT ACCTCTCTTC AAGGGGCAAT GTTCAAAGTC ATGAAAGAAG 20460

AAAGCGGACA CTATACTCCT GTTCTTCAAA ATGGTAAGGA AGTAGTTGTA ACATCAGGGA 20520

AAGATGGTCG TTTCCGAGTG GAAGGTCTAG AGTATGGGAC ATACTATTTA TGGGAGCTCC 20580

AAGCTCCAAC TGGTTATGTT CAATTAACAT CGCCTGTTTC CTTTACAATC GGGAAAGATA 20640

CTCGTAAGGA ACTGGTAACA GTGGTTAAAA ATAACAAGCG ACCACGGATT GATGTGCCAG 20700

ATACAGGGGA AGAAACCCTT GTATATCTTG ATGCTTGTTG CCATTTTGTT GTTTGGTAGT 20760

GGTTATTGTC TTACGAAAAA ACCAAATAAC TGATATTCAA TGTACATCAT TATGAATAGG 20820

ATAGCAGGCT GAAGGGAAGA CCAGAGTACT CTGAGGTGAT GTTAATCAGG AATCATGGTG 20880

ATGTGGCATG AATCATCAAT AACGGATATG AGGCTGGGCA GATTGTGCCA GCCTCATTGT 20940

GGGTTATTGT TTGTAAAACG ATAGGACTGG TCTGGTAATC ATTTTA 20986

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21040 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

CCCAGCAAAA AGCCATCCGA AGATGACTTT TTTGCTATTT AATTTCTGTA TAAGTTACTT 60

CCAAGCCACG CTTAACAGCT GGACGATTGG CAATTTTTTC TGCCCATTTT ACTAGATTTT 120

GATAACTTGA GGCATCCAAG AATTTTGCAG AACCTTGGTA AAGATTTCCT TGAACTAACT 180

GTCCATACCA AGACCAGATA GCAATATCTG CAATCGTATA GTCATTGCCT GCAATATAAG 240

GTTTCTGAGC CAATTCCTTA TCCAATAAAT CCAACTGGCG TTTCACTTCC ATCGTAAAAC 300

GGTTAATAGG ATATTCCAAT TTTTCAGGAG CATAATTGAA GAAATGTCCA AATCCCCCAC 360

CTAGAAAAGG TGCTGCACCT GCTTGCCAGA ATAGCCAATT CAAAACTTCT ACCTTTTCCA 420

CAGGATTACT TGGTAAAAAG GCTCCAAATT TCTCAGCAAG GTAAAGAAGA ATATGAGCAG 480

ACTCAAAGAC TCTTACGTTT TCAGTACCTG ACTGGTCCAA TAAGGCTGGA ATCTTGGAAT 540

TTGGATTGAG CTTCACAAAG TCTGATCCGA ATTGATCCCC ATCCATGATA GCAATCTTAT 600

ACAAGTCGTA AGCCGCTTCC TTAAAACCAG CTTCTAGTAA TTCTTCCAAT AAGATAGTAA 660

CCTTCACACC ATTTGGTGTT CCCAGTGAAT AAAGCTGAAA AGCTTGTTCT CCTTTTGGCA 720

AGTTTTGTTC GAAACGGGCA CCTGCTGTTG GTCTGTTTAG CCCCGTAAAA GCTCCTTGAT 780

TACTAGCTTC ATCCTGCCAT ACGGTCGGTA ATTGATATGC TGACATCCGA AACCTCCCTT 840

AAATCGCATT CTTGTCAAAA CCGAGTTTGC GTTGAATAAA CTTAACGATT TCGACGATGA 900

TAATCATTGA GAAGCTTCCA GCCATAACAA TTCCCCATTG TGACAAGTCT AGTTTGGTTA 960

CGTGGAAGAT TCCTTCAAGC GGTTCTACAA CGATTGTTGC CATGAGAAGG ATAAAGGATA 1020

CCAAGATGGA CCAGTTAAAG GTCTTAGACT TGAATGGGCC AACTGTCAAG ATGGATTGGT 1080

AGACAGACTT GACATTGTAG GCATGGAAGA GCTGAATCAA ACCAAGGGTT GCAAAGGCCA 1140

TCGTTAGGGC ATCTGCATGA ATAGCATGAT TGTCACCCAC ATGAACTGGG TAAGCAATCG 1200

CAAGGCCATA AACACTCATA ACAAGAGCTG CTTGGAGTAC ACCTTGATAA ATGATAGAAC 1260

TCAAAACACC ACCTGAGAAG AAGCTTGCCT TGCGTCCACG TGGTTTATGA TTCATGACAC 1320

CAGGTTCCGC AGGTTCAACA CCAAGAGCGA TAGCTGGGAA GGTATCCGTT ACCAAGTTGA 1380

TCCACAAAAG ATGAACCGGC TGTAAGACAT CCCAACCAAA CAAGGTTGAT AGGAAGATGG 1440

TTAATACTTC AGCAGTATTA GCAGAAAGTA GGTACTGAAT AGTCTTTTGA ATGTTTGAGA 1500

AGACCTTACG TCCTTCTTCC ACTGCGACGA TAATAGTCGC AAAGTTATCA TCTGCAAGAA 1560

TCATATCAGA AGCCCCCTTA GAAACCTCTG TACCAGTGAT TCCCATACCG ATACCGATAT 1620

CGGCTGTTTT CAGAGCTGGC GCGTCATTGA CACCGTCACC TGTCATGGCA ACGACTTTAC 1680

CTTGTTTTTG CCAAGCCTTG ACGATACGAA CCTTGTGTTC TGGAGACACA CGGGCATAAA 1740

CAGAGTATTG ACCAACGACT TTTTCAAATT CTTCATCTGA CAGTTCATTG AGTTCAGCAC 1800

CAGTTAAAAC GTGACCTTCT GTATCGTTTG CGTCAATGAT TCCCAAACGT TTGGCAATGG 1860

CTTCCGCTGT GTCTTGGTGG TCACCTGTAA TCATAATTGG ACGGATTCCC GCTTCCTTAG 1920

CCACACGAAC AGCCTCAGCG GCTTCAGGAC GTTCAGGGTC AATCATCCCA ATCAAACCAG 1980

TAAAAATTAA ATCATTTTCA AGCTCTTCAG AAGTGAGATT TTCTGGAATA CTATCGATAA 2040

TCTTATAAGC ACCTGCAAGG ACACGCAAGG CTTGATCAGC CATTTCAGAA TTGTTTGTAC 2100

GAATGAGATT TGTAACCTTC TCATCAATCG GAGCAATATC CCCAGCCTTA TCACGAAGAA 2160

GACAACGTTT TAAGAGTTGG TCTGGCGCAC CCTTGACTGC TACAAGGAAA CGACCATCTG 2220

GCAATGGGTG AACTGTTGAC ATGAGCTTAC GGTCAGAGTC AAATGGCAAT TCAGCTACAC 2280

GAGGATATTT CTCTAAGAAA CCTTTGACAT CATAGCCCTT GTCCAAGGCA TATTGGATAA 2340

AGGCTGTTTC GGTTGGGTCA CCAATCAAGT TACCTTCCAC ATCGATTTTC GTATCATTGG 2400

CCAAGACAAC TGAACGAAGT AGTGGCATTT CAAGACCTAG TTCAATATCA TCAGCTGAGT 2460

CATGTAGAAC CGCATCGTAG AAGACTTTTT CGACTGTCAT CTTGTTCATA GTCAGCGTAC 2520

CAGTCTTATC AGAAGCGATG ATTTCAGTTG AACCAAGTGT TTCAACTGCT GGCAACTTAC 2580

GAACGATGGA ATGTCGTTTG GCCAAAACTT GAGTACCAAG AGAAAGAACG ATGGTAACGA 2640

TAGCAGGAAG TCCTTCTGGA ATGGCTGCAA CGGCAAGGGC AACAGAAGTC AACAACTCAC 2700

CAAGTGGATT TTTCCCTTGA ATGAAGACAC CCACTACAAA AGTAACAAGG GCAATGACCA 2760

AGATAGCATA GGTCAAGACC TTAGAAAGGT TGTTCAAATT TTGTTTGAGT GGTGTATCAG 2820

TCTCATCCGC ATCTTGAAGC ATACCAGCAA TATGACCAAC TTCAGTGTAC ATACCTGTAT 2880

TGACAACAAC ACCCATCCCA CGACCATAGG TTACGTTTGA GTTTTGGAAG GCCATGTTGA 2940

CACGGTCACC AATACCAGCA TCTGTCGCAA GCTCGACTGA CAAGTCTTTT TCGACTGGTA 3000

CAGATTCACC TGTCAAGGCT GCTTCTTCAA TTTTAAGAGA GTTGGCTTCT ATCAAACGTA 3060

GGTCCGCTGG TACCACGTCA CCTGCTTCAA GGGCAACGAT ATCGCCTGGT ACCAATTCTT 3120

TAGAGTCAAT CTCTGCCATG TGTCCATCAC GAAGAACGCG GGCAACTGGA CTAGACATGG 3180

ATTTGAGGGC TTCAATAGCT TCTTCAGCTT TTCCTTCTTG GTAAACACCA AAGGCAGCGT 3240

TGATGATAAC CACAGCTAGG ATGATAATGG CATCTGCGAT ATCTTCCCCA CCAGAAGTCA 3300

CGACTGACAA GATTGctGCC GCAACTAGGA TGATAATCAT CAAATCCTTA AATTGCTCGA 3360

TGAATTTGAC CAAGATTGAT CGTTTCTCGC CTTCTTCGAG TTCATTGTGC CCAAATTCGG 3420

CAAGGCGCTT TTCCGCCTCA CTTGATGACA AACCTTGCTC GGTCGCATCC ACAGCCTGCA 3480

AGACCTCTTC AGGGCTCTGA GTATAAAACG CTTGGCGTTT TTGTTCTTTT GACATGTGTC 3540

TCCTCCTTGA CATTGTGTGC AAAACAGACT CTCTTTCTGT CATAGCTTTT CACGACAAAC 3600

AAAAAGAAAC CTGTTAATCA TAACAAGTCT CGCTGTTTAA GATAGGGCCG GAAAGCATAC 3660

TTTTCAGCAT AAAATTCGGA ATGACGACAC TATCACAGGT TTCTGCCAGC TACTCCCTTG 3720

AGTAGTACCA TTATACCAAA TTTTGGGGAG TTTTCAAAGA GTAAAAACTG CCTTATTTGA 3780

ATTTTTCCTT GAAAACCAGT ATAATGGTAG AATGCTATGT GACTAGAAAG GAAGTTGAAT 3840

GAAGCAATCT ATCTCAAATC TCAAGTTAGC TGAGCGTGGA GCCATTATCA GTATTTCGAC 3900

CTATTTGATC TTGTCTGCAG CCAAATTAGC AGCTGGTCAT CTCCtTCATT CATCCAGTTT 3960

GGTGGCCGAT GGTTTTAATA ACGTATCGGA CATCATTGGA AATGTGGCCC TCTTAATCGG 4020

GATTCGGATG GCGCGCCACC TGCAGACCGT GACCACCGTT TTGGTCATTG GAAGATTGAA 4080

GATTTGGCAA GCTTGATCAC TTCTATCATC ATGTTCTATG TCGGTTTCGA TGTTCTAAGA 4140

GATACCATTC AAAAGATTCT CAGTCGGGAA GAAACGGTCA TTGATCCTCT TGGTGCAACT 4200

CTAGGAATCA TTTCTGCAGC GATTATGTTT GTGGTCTATC TCTACAATAC TCGCCTCAGT 4260

AAGAAATCCA ACTCCAATGC GCTGAAGGCA GCTGCTAAGG ACAATCTTTC TGACGCTGTT 4320

ACCTCACTTG GAACCGCCAT TGCCATCCTA GCTAGTAGTT TCAATTATCC GATTGTGGAT 4380

AAACTGGTTG CTATCATCAT CACTTTCTTT ATCTTGAAGA CTGCCTATGA TATCTTCATC 4440

GAGTCTTCCT TTAGTCTTTC AGATGGCTTT GACGACCGCC TGCTCGAGGA CTACCAAAAG 4500

GCTATCATGG AAATTCCCAA AATCAGCAAG GTCAAATCGC AAAGAGGTCG CACCTACGGT 4560

AGCAACATCT ACCTGGATAT TACACTAGAG ATGAATCCTG ACTTGTCTGT TTTTGAAAGC 4620

CATGAAATCG CGGATCAGGT CGAGTCTATG CTGGAGGAGC GTTTTGGCGT CTTTGATACC 4680

GATGTCCATA TCGAACCAGC ACCTATCCCT GAGGATGAAA TTTTAGACAA TGTCTATAAA 4740

AAATTGCTTA TGCGTGAACA ATTGATTGAC CAAGGAAACC AACTAGAAGA ACTCTTGACT 4800

GATGATTTTG TCTATATTCG CCAAGATGGA GAGCAGATGG ATAAAGAGGC TTATAAGACC 4860

AAAAAAGAGT TAAATTCTGC TATCAAGGAC ATTCAAATTA CTTCCATCAG TCAAAAAACC 4920

AAACTCATCT GCTATGAGTT AGATGGTATC ATCCATACCA GTATCTGGCG TCGCCACGAA 4980

ACCTGGCAAA ATATCTTTCA TCAAGAAACC AAAAAAGAAT AGAGAAATCC TTTCATGAGA 5040

CGGGATTTTT CTATTCTTTT ATACTCAATA AAAATCAAAG TGCAAATTAG GAAGCCGGTC 5100

ACAGGCTGTA CTTGAGTCGG CAATGTGAAG CCGACATAGT TTGCACTTTG ATTTTCGAAT 5160

AGTCTTAACT ATCAAATTCA CTGAGATACT CATAGCGTTC GTATTTTTCA AGGAGTGCTT 5220

CATTTTTCTC ATCCAATTCT TTTTGGAGAG TAGCCAGCTT ACCAAAGTCA GAGCCGTTAG 5280

CCTGCATTTC CTCTTCAATA GCAGCGATAC GTTTTTCCAA GGTTTCAATA TCACCTTCAA 5340

TACTTGCCCA CTCCTGCTTT TCTTGGTAGG TCATGCGTTT CTTGTCTTCT CGAACCTTGA 5400

CCACTTTTTC CTTTTCGGCC TTTTGCACTT GATTGGCCAT ATCTGTTTCA AAAGCTTTTT 5460

CATCAAGATA GTCGGTGTAA TGACCAAAGA AAGGACGAAT CTTGCCATCC TCAAAAGCGA 5520

GAATCTTGGT CGCTACCTTA TCCAAGAAAT AGCGGTCGTG ACTGACTGTT AAAACGGGAC 5580

CTGCAAAACC TTGCAAGAAA TTCTCTAAGA CTGTCAAAGT TGCAATATCT AGGTCATTGG 5640

TTGGCTCGTC TAAAAGAAGA ACATTTGGTT TTTCCAAAAG CAGTTTGAGG AGATAAAGAC 5700

GTTTTTTCTC ACCCCCTGAC AATTTCTCAA TCAAAGTCCC ATGCGTCGAA CGTGGGAAGA 5760

GGAATTGCTC CAGCAACTCA GCGATGGAAG TCGTAGAACC ACCACTGGTC TTGACCTCCT 5820

CTGCCACTTC CTGCAGGTAA TTGATCACAC GCTTGCTTTC ATCCAAACCC TCAATTTGTT 5880

GAGAGAAATA GGCGATGCGA ACAGTTTCCC CAATCACAAC TTGTCCTGCT GTCGGCTCAA 5940

GACTTCCTGC AATCAGGTTA AGTAGGGTTG ATTTTCCAAC ACCATTGTCC CCAACAATTC 6000

CAATACGGTC TTTAGCCTGA ACTAAGAGAT TAAAATTTTG CAAAATGGGC TTATTTTCAT 6060

AGGCAAAGGA AACATCCTGA AACTCGATGA CTTTCTTCCC AATCCGACTG GTTTCAAAGT 6120

TCATAGTCAA GTCTGTCTCA GCACTACTGC CTGAAACTTC CTTTTTCAGA TCATGGAAAC 6180

GATTGATACG AGCTTGTTGC TTGGTCGCAC GCGCCTGCGG TTGTCTGCGC ATCCAGGCCA 6240

ATTCTTGTTT GTAGAGTTGT TCTTTTTTGT GAAGAAGAGC CGCGTCGCGC TCATCCTGTT 6300

CCGCCTTTAG GCGAACATAG TCCTGGTAAT TTCCCTGGTA CTCGGTCAAG CCTGCACGAT 6360

CCAACTCGAA AATCCGTGTT GACAAAGCGT CTAAGAAATA ACGATCGTGA GTGATAAAAA 6420

GGACGGTCTT CTTAGAATTT TTCAAAAAGA GGGTCAGCCA CTCAATAATC GCAATATCCA 6480

GATGGTTGGT CGGCTCATCC AAAAGCAAGA GGTCGTGGTT GCCAAGTAAG ACTTGTGCCA 6540

ACTGTACCCG TCTTCTCAGA CCACCTGACA ATTCCCCAAC AGGAGTAGAT AAGTCTTGAA 6600

TGCCCAATTT GCTAAGAACG GTCTTGACCT GACTTTCGAT TTCCCAAGCT TGGAGAGAGT 6660

CCATCTCTGC CATGACACGT TCCAAACGCG CCTGCTTGTC CTCACTATAG TCGAGCATAA 6720

TCAATTCATA CTCACGAATG AGCTGGATTT CCTTGAGTTC ACTAGATAGA ACCGTATCCA 6780

AAACTGTCTT TCTATCATCA AAATCAGGAT CCTGAGTCAA GTAACCAATC TGGTAATCAT 6840

TTTTAGCTGA AAAAGGACTG ACATCCCCAT CAAATCCAGA AACACCAGAA AGGACGTCCA 6900

AAAGGGTGGT CTTGCCAGTC CCATTGACAC CGATTAAACC AATTCTGTCT AAGTCATGGA 6960

TAATAAAGGA AATATCCCTA AAAACGGTCT TGTCACCAAC GGATTTACTT AGTTTTTCAA 7020

CGATAAAATC ACTCATTTTT TCTCCCTCAG GTAAGCATGG ATGGCTTCAC GATTATTCTC 7080

CAATTCTCCA TCGACAATGG CAAACTCAAT CTCTGTTAAA ATCTCTCCCA AGTCTGGGCC 7140

TGGCTGATAG CCATATTCCT TGATCAAAAT ACCGCCATTA ATCTGAATCT CTTTCTTGTC 7200

ATGGATAGTC AAGCTTTGGT ATTTTTCTGT GATGGCTTGT GGGTTGACTT CTTTTCCTTG 7260

AGCTTGACGA AGATTTTCAG CCTGTAAAAG CAAATCTATG TCAAAGCGAT AACAATCTCG 7320

CTTGCTCAAT TCTCCATTTT CACGCAGAGC CAAAATAATC AGCAAATCCT GAACTTGCTT 7380

GGCAAACTGG CGTGAGGTCT TCCAAGATTT CAAAAATGAC TGCGCATTTT CAATCTCCAA 7440

AGCCCATAGT AAAGCCGCCC AGGCTTGTTC AGAGGATTCA AAAGTAAAAT CAGTCTCCAA 7500

ATCAAACAGT CTGTTGAGCT TGTCCTGGCT AGATGCCATA TCAGGGAGAT AGTCATAAGC 7560

TTGACTCTCA ATCATGGAAG CCAAGCCCCT TCTCCAAAAT GGAGCCAGCA AGAGTTTATC 7620

AAACTCGACG AAGGTACGCT CTACAGAAAT TTTCTCCAAA AGCGGCGTCA AGGTCTTCAT 7680

AGCTTTAAAT GTTTCTGGCT CAAGTGCAAA ACCAAGACTA GCCTGAAAAC GGAAACCACG 7740

CATAATCCGT AAAGCATCTT CGTTGAAACG CTCACTAGCC ACTCCAACTG CTCGCAAGAC 7800

TTGCTTTTCC AAATCTTCTA AACCATGGAA CAAGTCAACG ATTTCTCCTG TCTCATCCAA 7860

GGCAAAGGCG TTGACTGTGA AATCACGGCG TTTGAGGTCT TCTTCTAGCG ATCGTACAAA 7920

GGAAACCGCA CTGGGTCTGC GATAGTCCAC ATAGACATCC TCTGTCCGAA AGGTTGTTAC 7980

CTCATACTCC TCATCCCCAT CTAAGACCAA GACGGTTCCA TGCTCGATTC CGATATCGGC 8040

TGTTCGCGGA AAAATCTGCT TGGTCTCTTC TGGATAAGAA GACGTCGCAA TATCCACATC 8100

GTGGATAGGG CTATGGAGAA GGGCATCTCG AACAGAGCCC CCAACAAAAT AAGCCTCAAA 8160

GCCTGCTTCT TTAATTTTTT CTAATACTGG TAAAGCCTTC TGAAATTCAG AAGGCATTTG 8220

CGTTAATCTC ATAATAAGTG TTCTAATCCA TAGACAAGCT CATGACGCTT GACAACTTCT 8280

TTAATTCCCA AATTGACTCC TGTCATGAAG GAGATGCGAT CATAGGAGTC ATGACGGAGG 8340

GTCAACCCTT CTCCCTGATT GCCAAAGATG ACTTCCTGAT GAGCTACCAA GCCTGGCAAA 8400

CGAACTGAGT GGATGCGCAT ACCATCAAAG TCAGCACCAC GAGCACCAGC AATCAGCTCT 8460

TCCTCATCTG CTGCACCTTG CTGAATTGAC TCTCGAACCT CTGCCATCAA CTCAGCTGTT 8520

TTAATGGCTG TTCCACTCGG AGCATCCTTT TTCTTGTCAT GATGGAGCTC AATAATCTCC 8580

ACATTTGGGA AATATTTGGC AGCCTGCGTC GCAAATTGCA TGAGTAAGAC AGCACCCAAG 8640

GCAAAGTTAG GGGCAATCAG GCCACCCAAG TCTTGGGCAC GAGAAAATTC TTTTAGCTCT 8700

GCAATTTCTT CACTCGTGAA ACCAGTCGTT CCAACTACTG GAGCAAAGCC ATTTTCAAGA 8760

GCAAAACGTG TATTTTCGTA GGCAACAGCT GGAGTAGTAA AATCTACCCA GACATCCGCT 8820

TCAAAACCAG CTAAATCAGC CTTATCCTTG AAAACAGGAA TACCCTGCCA TTCTGACTCA 8880

GACTCAAAAG GATCCAAAAC TGCCACCAAG TCCAAGTCTG GATCAGTCAA TACCATCTGA 8940

CAAGCAGCCT GGCCCATCTT TCCCTTAAAA CCGGCAATAA TTACTCGAAT ACTCATCTCT 9000

ACTCCTGTCT AAGATACAAA GTCCGTAAGA ACACAAAGTG AAAATAGGAA TTCCAATCAA 9060

GAAGTGTCTA CTTCTTGGAA GAACTATCTT TTTCACACAG GGTTCCAGGC GTGTTCAATT 9120

ATCAAGATAC AAAGGACCTT AGCTGCCTCT GAAAAATAGG GAATGGCACT GACTTTCCAC 9180

GAAAGGCAAG ACAGGCATCT TTTTTCAAGA GGCAGGTAGT CCGTGTTCAA TTTCTAAGAT 9240

ACAAGGCATC TTAACTAGCC TAGAAGCGCC AACTAAATCA CTGGAATATA ACCCAGAGCA 9300

ATACTTCCTG CTCCTAGGTG CGTTCCAATG ACACTACCAA ATGTAGCAAG TGAAACATCC 9360

GAACCCAAGC CAAAATCAAG CAAGTGCTGA CGCAATTCTT CAGCCTTTTC AGGAGCATTC 9420

CCATGAATGA CAATGACCCG GTATTGACCT GAAGCCGTTG TTTCCTTGAT AATTTCAATT 9480

AAGCGCTTGG TGGCCTTCTT TTCAGTACGA ACTTTTTCGT AAACTTCAAT CACACCTTGA 9540

TCGTTAAAAT AAAGGATTGG CTTAATGCTA AGCAAATTGC CCAAAATGGC AGCCCCATTT 9600

GAAAGGCGTC CACCTTTTAC CAAATGATCC AAGTCATCTA CCATGATAAA GGCTGACGTA 9660

CGGCTGATTT GAATGGCTAG CTTATCCTGA ATGCTGGCAA AATCATCGCC CTGATCACGC 9720

CAATTAAAGA CGCTTTCAAC CATGATGCCT AGGGGAGCAC TTGTAATCAA AGTGTCTGGG 9780

AAAGCAATGG TTAAGCCCTC ATAGTCATCG ACCATATACT GGATATTTTG GTAAAAACCT 9840

GAAATTCCAG AAGATAGGAA AAGCCCCAAG GCATGTGTAT AGCCTTGTTC TTTGAGCGAA 9900

GTTAAGATCT CATCTAACTT GGCAATACTT GGTTGACTGG TCTTAGGCAA TTCAGAAGCC 9960

TGAGCCATTT TTTGGTAAAA TTCCTCAGCA GACAGATTGA TGCCTTCGAC ATATTCCTCA 10020

CCATCAATAT TGACAGGAAT ATCCAAGACA AACAAGTCTT CTCTTTGCAA GATCTCTGCA 10080

CTGAGATAAG CAGAGGAATC TGTGAAAACA GCTAATTTCA TATTAGAACT CCAAATTAAT 10140

TCCTGGTAAG TCTAATGCAA TTTCAGTCAC TTCGTAAGTC AAACGATTGA GCATGTTCAA 10200

ACATGGACGA GCCAAGGTTT CCACCTCTTC TTGGTTCAAT TCACTTGGTT CATTGACAAT 10260

ACGGCCATCG ATATGGTTTA CTTGTGAGAT TGTTCCACTA ATGACAAACT TATCAAATAC 10320

AATCATAAAG CTCAAGATGA CAATCAAGGA AGTCACTTGA TTTTCTTGGT CATGTTGGAG 10380

CAATTGGAAA TTCACATCCA CCTTGGTTTC AGGAGCTCCA TTTTCATTTT CCCATTCAAA 10440

ATTACGCGCA TCAAAATGAT ACTGACTAAC AAATTCTTGT TCACGTTTAA GATTCATGTC 10500

TTTCTCCATC GGCTACAATA TTATAAGCTA TTGTACCATA ATTTTTTATT TTCATCTAGT 10560

TTTCTAGGAT TTAGTCAATC CCAATTTCAG CACGAACTAC ATCTGTGATG GTATCAACAT 10620

AGTAGTTTAC TTCTTCTGTT GTAGGCGCTT CTGCCATAAC ACGCAAGAGG GGTTCTGTTC 10680

CACTTGGACG AACAAGGATA CGGCCGTTCC CCGCCATTTC TTCTTCCATC TTCTCGATGA 10740

TGGCCTTGAT AGCTGGCACT TCCATGGCCT TTTCCTTCAT GACGTTTTCC ACTCGGATAT 10800

TAACTAATTT TTGTGGATAA ATCGTTACTT CTGCCGCCAA CTCTGATAAG CTCTTACCAG 10860

TTTCCTTCAT GATTTTAGTC AATTGAACTG CTGATAATTG ACCATCACCT GTGGTATTGT 10920

AATCCATCAA GATAACGTGA CCAGACTGTT CACCACCAAG GTTGTAGCCT GATTTTCTCA 10980

TTTCTTCAAC AACGTAGCGG TCACCAACTG CAGTAACTGC CTTGTTAATA CCTTCGCGAT 11040

TCAAGGCCTT GTGGAAACCA AGGTTAGACA TAACAGTTGT CACAATTGTA TTTTGAGCCA 11100

ATTGTCCTTT TTCAGAAAGG TATTTTCCGA TGATGTACAT AATCTTGTCA CCATCAACGA 11160

TGTCACCATT CTCATCAACA GCAATCAAGC GGTCACTGTC TCCATCAAAG GCCAAACCAA 11220

TAGCTGACCC ACTTTCTTTG ACCACTTCTT GAAGGGCTTC TGGATGTGTT GAACCAACAT 11280

TAAGGTTGAT GTTAAGACCG TCTGGTGTTT CCCCGATAAC CGTCAATTGG GCACCAAGGT 11340

CTGCAAAGAT TTGACGGGCA CTGGTAGAAG CTGCTCCATT AGCTGTATCC AAGGCAACCT 11400

TCATTCCATC AAGAGGAGTT CCAGTTGAAA CAAGGTATCC TTCATACTTA CGCArGCttC 11460

TGGATAATCT ACCAAAATTC CTAAGCCTTC TGCACTTGGA CGAGGAAGAG TGTCTTCCTC 11520

AGCATCTAGC AAGGCTTCAA TTTCTGCTTC TTTTTCATCA TCTAGTTTGA AGCCATCACC 11580

GCCAAAGAAC TTGATTCCGT TATCAAGGGC TGGGTTGTGG CTAGCAGAAA TCATGACACC 11640

GGCACTTGCT CCTTCAGTTT CAACCAAGTA AGCTACTGCT GGTGTTGCAA GGACACCAAG 11700

TTTGTATACG TGAATCCCTA CTGAAAgGAG ACCTGCCACC AAGGCCGATT CCAACATTTC 11760

CCCTGAAATA CGTGTGTCAC GTCCTACAAA GACTTTCGGC GCTTCCGTTT CATGTTGACT 11820

AAGAACATAG CCTCCAAAAC GTCCTAGTTT AAAGGCTAAT TCTGGTGTTA GTTCTAGGTT 11880

AGCTTCTCCA CGGACTCCAT CAGTCCCAAA ATATTTACCC ATTGTTATAA AATCCTTTTC 11940

TATTTTTAAT TCGTTTTTGA ACTAGTTGCT TTCGTTGACG AAGATGTCTC CGATGAACTG 12000

CTTGTACTTG AATTTGATGT GCTTGAACTT GGTGCTACTG GTTTTGTAGT CACCTTCATT 12060

ATTGTATCAA ACGGAGTGAT AACTGCCGGT AAGACAACAC CATTGCGGTC GATTGCCTGC 12120

AAAGGTACTG AACCACTGTA ATTACCTGTT ATACGTTCGC TAGTTGGCAA AACAGCGATA 12180

ATCTTATCAA TTCTATCCAA TGTCTCTTGG TCACTCGTAA TAGACACTTC TTTATCTGAC 12240

ACCATGACAT TTTCAATTTG TACCCGACTA TCAATTTGAC TAGGGTCAAT CTCTGGTACA 12300

ATCTTTACCT TATCCTTCTG AGCCTTCTTA CCAATCTTGA CTGTAATTTT TTGCGGAGTC 12360

GCCACAGCGG TCAGCCCATT GGGTAAATCT TCAATGCTCA AAGGAACTTC AATCGTTCCA 12420

ACACCGGCAT CTGTTAGGTC AGCAGTAACC TTGAATTTAC GTGTACTTTC TTGCATTTCA 12480

CTAGCTAGCG ATAGGCGATT TGCACCAGTC AAGACCACTG ATACTTCTGA AGCAAAACCG 12540

CTAATAAAAT ACTTATCACT ATTATAGCGT ATGTCAATAG GGACATTTGT TACTGTATTA 12600

GTATAGGTTT CCGTTTTTAC CTGCCTAGCA CTGGTACTGT TTTGAAAATT CGTCGCCGTA 12660

GCATAGACAA ATAAGACACA AGCAAAAAAG AGTGAGGATA TGATATATAA ACTATTTTTT 12720

TTCATGTTTC CATCCTCCTA GCAATCGTTC TTTAAAACTA AGACCCACTT CCTCTTTTGG 12780

AAGTAAGATT TCACGTAATT CTGTTTCAAA TTCATCAAGT GTTAGGTTGT GCTTAAACCT 12840

TCCATTATAG GTTATCGAAA TTCCTCCCGT TTCCTCTGAT ACGACAAAAG TCAAGGCATC 12900

TGAGACTTCT GATAAACCGA TAGCCGCCCG GTGTCTGGTC CCAAATTCCT TGGAAATCCC 12960

TGTGTTTTTT GTCAAGGGCA GATAGGCAGA CGTCACAGCG ATACGTTCTT CTTTGATAAT 13020

CACCGCACCA TCATGTAGGG GAGTGTTGGG AATAAAAATG TTAATGAGAA GTTCTGCAGA 13080

AATCTTAGCA TCCAAGGGAA TTCCTGTCGA AATATACTCC TGCAAGGTAC GTACACGCTG 13140

AATAGCAACC AAGGCCCCGA TTTTACGAGG ACTCATGTAT TCAACAGACT TAACAAAGGC 13200

ACGAATCATC TGTTCCTCAG CACTAATAGG GGCATTGGAA AAGAAATCTG TCGCTCTTCC 13260

CAAACGTTCC AAACCAGTCC GAATCTCTGG AGAGAAGATA ACAACCGCCG CAATAACCCC 13320

ATAAGTAATA ATTTGATTGA TTAACCAAGA AATCGTAGTC AAACCAATCA TATTTGCAAG 13380

GATTTGAGCT AAAATAAACA CCAAAACTCC ACGTACCAAA ATCATAATCT TGGTTCCTGC 13440

AATAGCTTTT GTAAAATGGT ATAAAATATA AGCAACAATC AAAATATCAA TCAGATTGAT 13500

AGCTATCGTC CATGGACTTG CAAACAAACT GGTCCAATAT TGCAGATTGG ATAATTGTTG 13560

AAAATTCATC CCTGATATCC TCCCTATCAA AACACTTTCG TCCTATTATA CCATTTTCTG 13620

GCATTTTTTT CCCTATCCTA GTCCATTTTA CATTGAACAA AAATATGATA AAATAAACTG 13680

ACTAAAAAAA ACAAAGGAGA AACTATGTCT CAACTCTATG ATATTACCAT TGTGGGTGGT 13740

GGTCCTGTCG GGCTTTTTGC AGCCTTTTAT GCCCACCTAC GCCAAGCCAA GGTTCAAATC 13800

ATCGACTCTC TTCCCCAGCT AGGTGGACAA CCTGCTATTC TCTACCCTGA AAAGGAAATC 13860

CTAGACGTAC CAGGCTTCCC AAACCTGACT GGAGAAGAGT TGACTAACCG CTTGATTGAA 13920

CAGCTAAATG GATTTGATAC CCCTATTCAT CTCAATGAAA CGGTTCTTGA GATTGACAAA 13980

CAAGAAGAAT TTGCCATCAC AACTTCTAAA GGAAGTCACC TGACTAAAAC AGTTATCATC 14040

GCTATGGGTG GCGGTGCCTT CAAACCACGT CCGCTGGAAC TTGAAGGGGT TGAGGGCTAT 14100

GAAAATATCC ACTACCACGT TTCTAACATT CAGCAATACG CTGGTAAGAA AGTGACGATT 14160

CTTGGTGGGG GAGACTCGGC TGTGGATTGG GCTTTGGCTT TTGAAAAAAT CGCACCAACT 14220

ACCCTTGTTC ACCGCAGAGA TAATTTCCGT GCCTTGGAAC ACAGTGTTCA AGCCTTGCAA 14280

GAATCATCTG TAACCATCAA GACACCATTC GCCCCTAGCC AACTCCTTGG AAATGGAAAA 14340

ACACTTGATA AACTTGAAAT CACAAAAGTC AAATCTGATG AAACTGAAAC CATTGACCTA 14400

GACCACCTCT TTGTCAACTA TGGTTTCAAA TCTTCTGTCG GTAACCTTAA AAACTGGGGG 14460

CTCGACCTCA ACCGTCACAA GATTATCGTC AACAGCAAAC AGGAATCCAG CCAAGCAGGT 14520

ATCTATGCTA TCGGTGACTG CTGCTACTAT GACGGAAAAA TTGATCTGAT TGCGACAGGC 14580

CTCGGAGAAG CTCCAACTGC TGTCAACAAC GCTATCAACT ACATTGACCC TGAACAAAAA 14640

GTACAACCAA AACACTCTAC TAGTTTATAA AAAAGAACCA CGAGTCACAT AGGATTCGTG 14700

GTTTTATAAT TCATCCGCTA TCTTATTGAT TTTTCTGAGT CTGTGATTGA CACCACTTTT 14760

GGTCAGAGGG GTGCTGAGAC TATCTGCTAA CTGCTGGATA GAGTAGTCTG GGTGCTGAAT 14820

CCTCAGTTGC GCCACTTCCT GCAAATCTAC TGGCAAATTT TCTAAGCCCA TGATATCTTT 14880

GATTTTACTG ATATTGTTAA TGGTCTTCAT GCTGGCAGAA ACTGTCCGAG CGATATTAGC 14940

TGTCTCGGCA TTATTAGCCC GATTGAGGTC GTTACGGGTT TCTCGCAAAA TCTTAACCCG 15000

CTCAAAATCA TCACGTGCCT GCATGGCTCC TATTACTATC AAGAAGTCCA TAATGTCTTC 15060

TGCTCGCTGG AGATAGGTCA CAGCCCCCTT CTTGCGCTCA AGCACCTTGG CATCCAGTAA 15120

AAACTGTTGG AGAAGGGAGG CAATTCCTTG CGCGTGGTCC AGATAAACAG AACTGATTTC 15180

CAACTGGTAC TTGCCTGACT CAGGGTCACG AATGCTCCCA TTTGCCAAGA AAGCGCCACA 15240

GAGATAGGCA CGACCTGCTT CCTCATCCGA TAAAATCGCC TCATCAATAC CTGTTTCCAG 15300

GCCAAAGAAA GAGTCTGCCA AGTGCAAATC ACTTAACAAA TCCTGCACCT TTTCATCTGT 15360

AAAAACGGTA TAGACGCGAT TCTTGCGAAG ATTGCTCCGT TGGTGGTGAC GAATTTCAGA 15420

TTTGATTTCA TAGAGATGGA GAAAGGACTC ATAGAGGTGA CGGGCCAGTT TGGCATTTTC 15480

TGTCACAACT GACAAAGTCA AGCCCGAAGT CGAGAGACCG ATGCTACCAG ACATTTTGAT 15540

AATGGCAGAT AATTCATGCC AGCTCAGATG GTGTTGGCCC AGGATTTCTT CTTTTACTGC 15600

TACTGTGAAA CTCATTTTTT CACCTGTATA ATGCGCATCA ACTCGTCCAC AATCAAATCT 15660

CCATCGTGGA AGGCACCGCC ATTTTCCAGA CGAAGGAAGT TAGATGAAAT CACGCGCGAA 15720

ACTTGCTTAC AAAGACCTAC AAAATCGTGT TCCACTTGCA CTAAGTATTC ATCAAAACGG 15780

TTGGAATTCA TGTATTCCTG AGGCACTTTT TCAATATTCA CCAAGACAGT GTCGATAAAA 15840

GGGCGACCAA GGTGACGATG CAAGACTTCC ACGTGGTCGC TATCTGTAAA GTGTTCCGTC 15900

TCCCCACGTT GGGTCATGAT ATTGCAGACA TAGGCAATTT CTGCCTTGGT TTCCAAAAGA 15960

GCCCGCCCAA TTTCCTTAAT CACGATATTG GGCAAAATAG AGGTAAAGAG GGAACCTGGC 16020

CCTAGGACAA TCATGTCACT TTCAAGGATG GTCTGCACTA CTCGACGGCT GGCCAGAGGC 16080

GTATCATCGT TTAGGGCATT GGTCACATAG ACATTGTCAA TTATGCCTCG ATGGTCTACA 16140

ATATGACTCT CTCCAGCCAC TTCTGTCCCA TCCTGAAAGA CTGCATGAAG GGTCAAAGGA 16200

TGGTCACTGG AAGGATAAAT TTTCCCTGTT GTATGGAAAA ATTTGCTCAA TAACTGCATG 16260

GCATTATAGG TTGAACCCTG CATTTCTGAC AAGCCAGCAA TGATGAGATT TCCCAATGGA 16320

TGGCCAGCAA AGGCTCCGGC ATCCTCAGAG AACCGATACT GAAAGACCTT CTCATAAAAC 16380

TTAGGCATAT CCGACATGGC CACAAGGACA TTACGAAGAT CACCTGGCGG TGTCAACTGT 16440

TGCATATTTT TTCGGAGTTC ACCTGAAGAA CCACCATCAT CTGCCACCGT CACGATAGCT 16500

GCGATTTCCA CATCTTTTTC CCGCAGACTT TTTAGAATGA CGGGACTTCC AGTCCCTCCA 16560

CCAATCACCG TTATCTTTGG TTTTCTCATG AACGGTTTAC CGTTTCCTTT CTGCGGTCTT 16620

TGTCGCGATG CCCTTCATTA ACAGACCAAT TCTTGGATAA GTCCTGCGCC AAGCGTTTAG 16680

CAAATGCCAC ACTACGGTGT TGTCCACCCG TACATCCCAT GGCAATGGTC AAAACGGACT 16740

TACCTTCCTT TTGGTAACTT GGCAGAATCG GCTCAATCAA GGCCAATAAA TGTTGATAAA 16800

AGTCTTCTGA CTCAGGATGG TTCATGACAT AATCATAAAC AGGTTCATCC ACACCCGTTT 16860

GGTTTCTCAG TTCTGGTAAA TAATAGGGAT TTGGCAAGAA ACGGACATCA AAGACCAAGT 16920

500

CCGCATCAAT CGGGATTCCA TACTTAAATC CGAAAGACAT GACTTCGATA CGGAAAGACT 16980

GGGCTTGTTC TTGGTCTGAA AACTGCTCTG CAAGGGTTTT GCGCAgCTCA CGTGGAGTGA 17040

GTTCAGTCGT ATCCACCACA TTTTGGCTCA TATTTTTCAA AGGTGCCAAG AGTTCACGTT 17100

CCAACTTGAT TCCATCTAAA ATACGACCGT CTGCTGCTAG TGGGTGACTC CGTCTGGTTT 17160

CCTTGTAACG AGCGACCAAT TCCTTATCAG CCGCATCCAA AAAGAGGATT TTGAAATCCA 17220

AACCATCTTG ATTTTCCAAC TCATCCAAAA CAGCTTGAAT CTCTGAAAAG AAAGAACGGC 17280

TACGCATATC CACTACCAAG GCCAACTTAG GATTGTCTTC CTTAATTTCA ACCAGCTGCA 17340

AAAACTTAGG CAAGAGAGCT GGCGGCATAT TATCAATGGT GAAATAACCT AGATCCTCGA 17400

AGGACTGAAT GGCTACAGTT TTCCCTGCGC CACTCATCCC TGTCACAATC ACCAAGTGAA 17460

GTTGTTTCTT TGTCATCTTT TTCTCCTTAT ATCAAAAGAA GTTTGGCAAC ACCAAACTTC 17520

AACTAGCTTA TCCAATCTCT GCGATGACTT CAATTTCGAC TTTTACATCA CGAGGAAGAC 17580

GAGCTACCTC CACAGCTGAA CGAGCTGGGA ATTCCTCTTT GAAGGCCGTT TGGTAAACCT 17640

CATTAAAAGG AACAAAGTCG TTCATATCGC TCAAGAAGCA AGTTGTTTTG ACAACATGGT 17700

CAAAGTCTGT TCCTGCTTCT GCCAAAATAG CACCGATGTT TTTCAAGACT TGCTCTGTCT 17760

GTTCTTGGAT ATTCTCTCCT ACAATTTCCC CAGTTTCAGG GGATAGGGGA ACTTGACCGC 17820

TAGCAAACAA AAGGTTGCCA ACGATTTTTC CTTGAACATA GGGTCCGATA GCCTTTGGGG 17880

CCTTATCTGT ATGAATTGTT TTTGCCATTT TCTTTTCCTC ACAATTTTTC TAAGATTGCA 17940

TCCCAAGCCT CATCCATCCC TGCCTTACTG ACAGATGAAA AGAGGATGAA ATCGTCACTC 18000

GGGTCAAAGT TTAATTTCTT TTTGATTGCT GATTCATGCT TGTTCCATTT ACCACGAGGA 18060

ATCTTGTCCG CCTTGGTCGC CACAATGATG ACTGGAATCT CATAATACTT GAGAAATTCG 18120

TACATCTGCA CATCATCTGC TGACGGGTCA TGACGAAGGT CAACTAGACT GACAACCGCA 18180

CGGAGATTTT CCCGAGTCGT TAAGTACTCC TCAATCATGC ACCCCCACTT TTCACGTTCC 18240

TTTTTAGAAA CACGAGCATA GCCATAACCA GGCACATCCA CAAAGCGCAT CTTGTCATCA 18300

ATGTTAAAAA AGTTCAGGAG CTGGGTTTTA CCAGGTTTTC CTGATGTACG GGCGAGATTC 18360

TTACGGTTCA ACATAGTGTT GATAAAGCTG GATTTACCAA CATTTGAACG CCCTGCTAGG 18420

GCAATCTCTG GCAGTTCATC CTGCGGATAG TGGGACTTAT TAGCTGCACT GAGCAAGATT 18480

TCAGCATTGT GTGTATTAAG TTCCATAGTC ACCTCTAGGC TGTTTCTAGG ATCGGTTTAT 18540

CCGTTCCATC TACAGTTTCT TTAGTGATGC GAACCAATTT CACATTTTCC TGACTCGGCA 18600

CCTCAAACAT GACATCTAGC ATGGTTTCTT CGATGATGGA GCGAAGTCCA CGCGCCCCTG 18660

TCTTCCGTTC GATTGCTTTA TTAGCAATCT CTTGAAGGGC TTCGTCGTCA AATTCCAACT 18720

CAACATCATC ATAAGAAAGC AAGGTTTGGT ATTGTTTCAC CAAGGCATTT CTTGGCTCTT 18780

TCAAGATGCG AACCAAGTCA TCAACGGTCA ATTGCTCAAG AGCCGCAAAA ACAGGCAAGC 18840

GTCCAATCAA CTCAGGGATA ATACCAAATT TTTGAATGTC TTCAGCGATG ATTTCTTGCA 18900

TGTATGAGCT GTTTTCGTCA ATCGCCTTAT TATTTTGACC AAATCCGATG ACTTTTTCAC 18960

CCAGACGTTG TTTGACAATT TCTTCAATAC CATCAAAAGC ACCACCCACG ATGAAGAGGA 19020

TATTTTTTGT ATCCACTTGA ATCATCTCTT GTTGTGGATG TTTGCGTCCA CCTTGAGGCG 19080

GTACGCTAGC AACAGTTCCC TCAATAATCT TGAGAAGGGC TTGTTGCACC CCTTCACCAG 19140

AAACATCACG TGTGATAGAC ACATTCTCAC TCTTCTTGGC AATCTTGTCA ATTTCATCCA 19200

CATAGATAAT GCCACGCTCT GCACGTTCGA TGTTAAAGTC AGCAACCTGC AAGAGTTTGA 19260

GGAGGATATT TTCCACATCC TCACCCACAT AACCAGCCTC CGTCAGAGCT GTCGCATCCG 19320

CAATAGCAAA AGGTACATTC AAGCTCTTAG CCAAGGTCTG GGCAAGGAAA GTTTTCCCTG 19380

AACCAGTTGG GCCAATCATC AAAATGTTTG ACTTCTGCAA ATCCACATCT TCTGACTCTT 19440

CGCGTGTATC GTGGAAATTG ATGCGTTTGT AGTGGTTATA AACCGCCACT GCCAAGGCAC 19500

GCTTGGCACG ATCTTGACCA ATTACATAGT GGTTCAAGAT ATGGAGGAGT TCAATTGGTT 19560

TTGGCACCTC AGACAAGTCT GCCAAGACTT CCTCAACCAA TTCTTCTCGA ATGATTTCCT 19620

GAGCTAACTC CACGCATTCA TTACAAATAA AAGCATTGTT GCCAGCAATT ATTTTTTGTA 19680

CTTCTTCTTG GTTTTTGCCA CAAAATGAGC AATAAACCAT CATATCATTT TTTCTATTTG 19740

TAGACATGAT TTCCTTCCAT TCTATACTGT CATTCTATCT AAAATAAGGT CATGTAAAAA 19800

GCATGAATAC TATTGACCAG ATTGGTAAAG GCATTTAACC AAAGGAGGAT AGAAAGCCCG 19860

TAACGCTTTT TACGAAAAGC TTGTGCTCCT GCCAGAAAGC AGATGAAACA CAGAAAAGCC 19920

GTGAATAGAC CAAATAAACT CCGTTCCATT AGACTTCCTT TCTCTTGCGG TATTGGATGG 19980

TAAAATCATA AGGATTCTTC TCATCTTTGG CGTAAAATTT GCTTGAAACT GTCTCAAAAA 20040

GAGACAAGTC AAGTTCTTCA GGGAAATAGG TATCTCCTTC CACCCGAGCA TGAATGTGAG 20100

TGACAATCAC TTCATCAAGG TAAGGTTCAA AAGCCTGAAA AATTTGCTTC CCACCGATAA 20160

TGTAGAGATT CTTTTCTTGA GCCTGATACC AGTCAAGAAC AGACTGGACG TCCTGAAAAG 20220

TAGCAACCCC ATCTATCTTT TCTTCCGGAT TACGCGTCAA AATCAAGGTT TCCCGTTTTG 20280

GAAGCAAGCG ACGCCCCATC CCATCAAAGG TCACACGCCC CATCAAGATA GCATGATTCA 20340

GAGTTGTTTC TTTAAAGTGC TGCAATTCTG CTGGCAAATG CCAAGGCAGA CGATTTTCCT 20400

TACCAATCAC ACCCTCTTCA TCCTGGGCCC AAATAGCTAC GATTTTCTTA GTCATGCTTC 20460

CATCCTTTTC ACTGATAGTA CTATTTTATC AAAAAACTCA AAAAAAGACT GGTTTGGAAT 20520

AGCTTACAAA ATAGAAAAAA TCTGTAAGAA ATTTCCTACA GATTTATCTA TGTTTCCTTA 20580

TTTCTTACAA ACCAGGTGCT TGTCCAAGTT CGGCTGCAAG CATCCAAATT GTTTTATCTG 20640

TTTCAGTTTT AGCGCCTGCA AAGATACCGT TTGTCACATC GTCACCTTCT TCATCAGTGA 20700

CATCCAAACC TTTTTGGAAA AGTTCTGACA AGTAACGGTA GATAACAAGA ACACGTTCCA 20760

AGCTTTCTTC AACATTACGG TATTCACCAG CTTCTTCTTC GATTTCACTA TTTTGAAGGA 20820

ACTCTGTCAA TGTAGAGAAT GGGCTTCCAC CGAGTGTAAT CAAGCGTTCA CTGATTTCAT 20880

CCAATTGACC GTCAAGAGCT TCCATGTACT CATCCATTTT TGGATGCCAT ACAAGGAAAC 20940

CACGACCATG CATATACCAG TGCACTTGGT GCAAAGCAAC GTGAGCTACA TACAAATCAG 21000

CAACAGCTTG GTTCAAGACT TCCTTTGTTT TTGCCAATGC 21040

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2387 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

ATTCTTAATA CGATTAAAAG GCTTATTACT AAAAGAAAAT TTCAGTTAGA TGAACTAAAC 60

TTGCTCGTCA AATCCCGATT TAACGAGATG TTTGGGGAAA ATAAAATATT TGAAAGCATT 120

GATAACTTAT TTGATATTAT AGATGGTGAT AGGGGCAAAA ATTATCCTAA ATCAGATGAG 180

TTGTTTAGTG AGGAGTACTG TTTATTTTTA AATACAAAGA ATGTTACTAA AAACGGATTT 240

TCATTCGATA CAAAGCAATT TATCACTAAA ACAAAGGATA AATTACTTCG AAAAGGCAAA 300

CTTGAGCGTT ATGATATAGT CTTGACAACA AGAGGTACTG TTGGAAATGT AGCGTACTAC 360

GATGAATTAA TAAAATATAA ACATTTACGT ATAAATTCAG GTATGGTAAT ATTACGTCCC 420

AAGACACCAA ATCTAAATCA GAAATTTATT ATCCATGTTT TAAGGAATAA TAATTATAGT 480

CGAGTGATAT CAGGAAGTGC TCAGCCTCAG TTACCAATTA CAAAATTAAA AAAAATACTT 540

CTCCCCCTCC CCCCACTAGC CCTCCAAAAT GAGTTCGCAG ACTTTGTAGT CCAGGTCGAC 600

AAATCACAAT TGGCAATCCA AAAATCTCTG GAAGAACTTG AAACTTTGAA GAAATCTCTG 660

ATGCAGGAGT ATTTTGGCTG ATATTCTGCC ATTGTAATTA CGGTAATGAT TTGTTATAAT 720

ACTTCAAAGG AGGAAATCAG ATGGTAGTAA AAACAAGAAA ACAAGGAAAT TCAATCACCA 780

TTACGATTCC AAGTGAATTT AATATTCCAA GTGGTGTTAA ATACGAAGCG AAATTGTTAC 840

CAAGTGGTGA GATTATCTTT ACTCCTGAAG AATTGGGGCA GCAGGTTTCT TATGTATCTG 900

ATGATGCCTT TGACTTAAAT TTAGATAAAA TATTTGACGA ATACGACGAT GTTTTCAAAG 960

CTTTGGTGGA AAAATGACAA TCTATTTGAC AGAAAAGCAA ATTGAAAAAA TAAATGCTTT 1020

AGCAATTCAA CGGTATTCTC CAAATGAGAA AATTCAAACA GTTAGTCCTT CTGCCTTAAA 1080

TATGATTGTG AACTTACCAG AACAATTTGT CTTTGGGAAG CCTCTTTATC CAACAATTTT 1140

TGATAAAGCA ACGATACTAT TTGTCCAATT GATAAAGAAG CATGTTTTTG CTAATGCTAA 1200

TAAAAGAACT GCTTTCTTCG TTTTGGTCAA ATTTTTACAA TTAAACGGCT ATCGTTTTTC 1260

TGTAACGGTA GAAGAAGCAG TAAAAATGTG TGTAACCATC GCAGTAGAAG CTTTAACTGA 1320

TGAAAAAATG ACAAGCTACT CCAAATGGAT TTCTGAACAT TCTGTTAGAG AAAAGGTCAA 1380

AAAGTAACCT AGTATGCTGG ATTTGAATGA GCACAAGAAA ATAAATGAAC AGACAATATT 1440

AGAATTCTGT AATGCAGAAA CTGATATTGT CTCTTTTTAT TGATGAATAA GAAAGTGAGA 1500

AATTATGGAA TCAAAAGTTA CAATTATCAT GCAAGAAATG TTACCTCTTT TAAATAATGA 1560

ACAATTACTA GCGTTGAGAG AGAGTTTAGA ACATCATCTA GTAGACGGAA AAAAGCAGCA 1620

GAAGTATTCG AATAATAACC TGTTGCAACT ATTTATTACC GCCAAGCAGG TAGAGGGCTG 1680

TAGCTCAAAA ACAATTCGTT ATTATCAGAG GACGATTGAA AACTTGTTTA ATGCTATTAA 1740

AGAGTCTGTG ACACAACTCA CAACAGATGA TTTAAGGAGT TATTTAGCAA ATTACCAGTC 1800

TGAAAAGGAT TGTAGTAAGG CAAATTTAGA CAATATTAGG CGTATATTGT CTTCTTTTTT 1860

TGCTTGGCTT GAGCAAGAGG ATATATCATT AAAATTCCCA TTCGACGGAT ACAGAAAATT 1920

AAGACTGAGC AAAATGTGAA GGAAACTTAT ACTGATGAAC ATTTGGAAAT TATGCGTGAT 1980

AACTGTGAAA ATTTGAGAGA TTTGGCAATA ATAGACCTAC TAGCATCGAC AGGTATGCGT 2040

GTAGGGGAGC TTGTACAGTT GAATCGTTCA GATATTGATT TTGAAAACAG AGAGTGTGTT 2100

GTCTTTGGTA AAGGAAAGAA GGAGAGACCA GTATATTTTG ACGCTCGTAC GAAAATTCAT 2160

TTAAGAAATT ATCTTAACGA CAGAAAAGAT AGTCACCCTG CTCTTTTTGT AACGCTAGTT 2220

GGAAAAGTCC AGAGGCTTGG AATTGCTGGT GTAGAGATTC GCTTAAGAAA GTTAGGAGAC 2280

AAACTCGGCA TACAAAAGGT TCACCCACAT AAGTTCAGAA GAACTTTAGC GACTAAGGCA 2340

ATTGATAAAG GTATGCCTAT CGAACAAGTC CAAAAACTGC TAGGTCA 2387

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10669 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

ATATTAAAGC GACTTTCTGT GCGCTAGGGA AAAATGTTCC TGGGAATGAG GACTTGGTGA 60

AGAGGATAAA ATCTGAAGGT CATGTTGTTG GAAACCATAG CTGGAGCCAT CCGATTCTCT 120

CGCAACTCTC TCTTGATGAA GCTAAAAAGC AGATTACTGA TACTGAGGAT GTGCTAACTA 180

AAGTGCTGGG TTCTAGTTCT AAACTCATGC GTCCACCTTA TGGTGCTATT ACAGATGATA 240

TTCGCAATAG CTTGGATTTG AGCTTTATCA TGTGGGATGT GGATAGTCTG GACTGGAAGA 300

GTAAAAATGA AGCATCTATT TTGACAGAAA TTCAGTATCA AGTAGCTAAT GGCTCTATCG 360

TTTTGATGCA TGATATTCAC AGTCCGACAG TCAATGCCTT GCCAAGGGTC ATTGAGTATT 420

TGAAAAATCA AGGTTATACC TTTGTGACCA TACCAGAGAT GCTCAATACT CGCCTAAAAG 480

CTCATGAGCT GTACTATAGT CGTGATGAAT AAGCAAGAAA AAATAGGTCT GTTAGATATT 540

TGACAGACTT ATTTTTTACA GAATATAGTA CTACTTAAAA AATGTTTTAT GCTATAATTG 600

ATGAATAAAA TAGAAGGAGA AGCATATGAA TACCTATCAA TTAAATAATG GAGTAGAAAT 660

TCCAGTATTG GGATTTGGAA CTTTTAAGGC TAAGGATGGA GAAGAAGCCT ATCGTGCAGT 720

GTTAGAAGCC TTGAAGGCTG GTTATCGTCA TATTGATACG GCGGCGATTT ATCAGAATGA 780

AGAAAGTGTT GGTCAAGCAA TCAAAGATAG CGGAGTTCCA CGTGAAGAAA TGTTCGTAAC 840

TACCAAGCTT TGGAATAGTC AGCAAACCTA TGAGCAAACT CGTCAAGCTT TGGAAAAATC 900

TATAGAAAAA CTGGGCTTGG ATTATTTGGA TTTGTATTTG ATTCATTGGC CGAACCCAAA 960

ACCGCTCAGA GAAAATGACG CATGGAAAAC TCGCAATGCG GAAGTTTGGA GAGCGATGGA 1020

AGACCTCTAT CAAGAAGGGA AAATCCGTGC TATCGGCGTT AGCAATTTTC TTCCCCATCA 1080

TTTGGATGCC TTGCTTGAAA CTGCAACTAT CGTTCCTGCG GTCAATCAAG TTCGCTTGGC 1140

GCCAGGTGTG TATCAAGATC AAGTCGTAGC TTACTGTCGT GAAAAGGGAA TTTTATTGGA 1200

AGCTTGGGGG CCTTTTGGAC AAGGAGAACT GTTTGATAGC AAGCAAGTCC AAGAAATAGC 1260

AGCAAATCAC GGAAAATCGG TTGCTCAGAT AGCCTTGGCC TGGAGCTTGG CAGAAGGATT 1320

TTTACCACTT CCAAAATCTG TCACAACCTC TCGTATTCAA GCTAATCTTG ATTGCTTTGG 1380

AATTGAACTG AGTCATGAGG AGAGAGAAAC CTTAAAAACG ATTGCTGTTC AATCGGGTGC 1440

TCCACGAGTT GATGATGTGG ATTTCTAGAA AATCATAAAA AGAATTGTAC ATTATTCTAA 1500

TTTTTGATAT AATAGTCAGC AGGAAAGAAA GTCTTATGGC GTTCTTCAAG CGAGCTTGGG 1560

ATAGTGGGAG CCAAGTAGGG CAAAATAAAG GGCTGGCGCT TTCTGTAGTA TTTTCAAAAA 1620

CAATGAAGTA ATAAATTAGG GTGGAACCGC GTTTCTGACG CCCCTAGGTT AAATCAACCT 1680

AGGATTGTCA GATGTGGTTC TTTTGCTTAT TCAGTCTATT GTGTGAAAGA AAGGAGAGCC 1740

GTGGACAACC TTTATCTTGT AAAAGACGAT AGTCAACTAG CTACATTTCG TGATTTTGTA 1800

GTAAGAAATA CTGAAAAGTT GAAAGATTAT CAATCTTTTT TAAAGAATGA ACTTGCAGTC 1860

TGTGATTTAC CGCAAGCTGT TATTTGGTCA GATTTTAATG CTGCTACACA GATTATTAGG 1920

GAAAGTGCTG TTCCAACCTA TACAAATAAT AGACGAGTGG TTATGACGCC TGATTTAGCT 1980

GTTTGGAAAG AATTGTATTT GTATCAGTTG ATGGACTACG AGTGTTCTGA GCAAACTCAA 2040

GCAATAGAAA GTCACTATCA TTCTTTATCT GAAAATTTCC TCTTACAGAT TGTAGGACAT 2100

GAGTTAGCTC ATTGGTCGGA CATTTTTTAG ATGATTTTGA TGGTTATGAC TCTTATATCT 2160

GGTTCGAAGA GGGGATGGTT GAATATATTA GTCGCAAGTA TTTCTTGACA GAAGAGGAAT 2220

TTCAAGCGGA AAAAATTTGT AATCAATCTC TCGTAGAACT TTTTCAGAAG AAGTATAGTT 2280

GGCATTCATT GAATGATTTT GGTTCTTCGA CTTATGATAA GAACTATGCA AGTATTTTTT 2340

ATGAATACTG GCGCAGCTTT TTGACAGTAG ATAAGTTGGT AGAAAATTTA GGTAGTGTAC 2400

AAGCGGTCTT AGATTCTTAT CATTTATGGG CAAATACAGA AAAAACTTTT CCCTTGTTAG 2460

ATTGGTTTGT TCAGCAGAAA TTAATTGAAA AAGAAATATA AAAACTAAAG GAGTAAACAA 2520

TGTCTAAGAA ATTAACATTT CACTGCATCA GTGGCAGAGA CCTCCTTACA GTCGGGCTGC 2580

TCCACGCTCA GCACTAGAGT GCCTGAGCTA GACGCAGTAC TAACTCGTCT TGCCTCGTAT 2640

GATCGACGAG GCAGACTCGT GTCGCAAGTA ATTATTTTTT ATTAAGGAGT ATTCAATGTC 2700

TAAGAAATTA ACATTTCACT GCGTCAGTGG CAGAAACCTC CTTACAGTCG GACTGCCCTA 2760

CGCTCAGCAC TAGAGTGCCT GAGCTAGACG CAGTACTAAC TCGTCTTGCC TCGTATAATC 2820

GACGAGGCAG ACTCGTGTCG CAAGAAATTA TTTTTTATTA AGGAGTATTC AATGTCTAAG 2880

AAATTAACAT TTCAAGAAAT TATTTTGACT TTGCAACAAT TTTGGAATGA CCAAGATTGT 2940

ATGCTTATGC AGGCTTATGA TAATGAAAAA GGTGCGGGGA CAATGAGTCC TTACACTTTC 3000

CTTCGTGCTA TCGGACCTGA GCCATGGAAT GCAGCTTATG TAGAGCCATC ACGTCGTCCT 3060

GCTGACGGTC GTTATGGGGA AAACCCTAAC CGTCTCTACC AACACCACCA ATTCCAGGTG 3120

GTCATGAAGC CTTCTCCATC AAATATCCAA GAACTTTACC TTGAGTCTTT GGAAAAATTG 3180

GGAATCAATC CTTTGGAGCA CGATATTCGT TTTGTTGAGG ACAACTGGGA AAACCCATCA 3240

ACTGGTTCAG CTGGTCTTGG TTGGGAAGTT TGGCTTGACG GAATGGAAAT CACTCAGTTC 3300

ACTTATTTCC AACAAGTCGG TGGATTGGCA ACTGGCCCTG TGACTGCGGA AGTTACCTAT 3360

GGTTTGGAGC GCTTGGCTTC TTACATTCAA GAAGTAGACT CTGTCTATGA TATCGAGTGG 3420

GCTGATGGTG TAAAATACGG AGAAATCTTT ATCCAGCCTG AGTATGAGCA CTCAAAATAT 3480

TCATTTGAAA TTTCGGACCA AGAAATGTTG CTTGAAAACT TTGATAAGTT TGAAAAAGAA 3540

GCTGGTCGTG CATTAGAAGA AGGCTTGGTA CACCCTGCCT ATGACTATGT TCTCAAATGT 3600

TCACATACCT TTAATCTGCT TGACGCGCGT GGTGCCGTAT CTGTAACAGA GCGTGCAGGC 3660

TATATCGCTC GTATCCGTAA CTTGGCCCGT GTCGTAGCCA AAACCTTTGT CGCAGAACGC 3720

AAACGCCTAG GCTACCCACT TTTGGATGAA GAAACAAGAG CTAAACTCCT AGCAGAAGAC 3780

GCAGAATAAA GAGAGTGACA AATTACGAAA ATGGGCGAAC AGAGTGAGCC CTGAGCCAGT 3840

TGCCGCAGTG ATGAAGGTAT CCTTAGTGAA ACTAAGGATA CTAGGCAAAA TTGGAGACTT 3900

TTGGCTCCAA TTTTAGCAAT GAAACAACGA AGTTGGTTGC TTGCGTGCCA ATCACATAAG 3960

GCAAACTGGA AAATAAAAAG ATACTTTTCG GAGAAAAAAC ATGACAAAAA ACTTATTAGT 4020

AGAACTCGGT CTTGAAGAAT TACCAGCCTA TGTTGTTACG CCAAGTGAAA AACAACTAGG 4080

CGAAAAAATG GCAGCCTTCC TCAAGGGAAA ACGCCTGTCT TTTGAAGCCA TTCAAACTTT 4140

CTCAACACCA CGTCGTTTGG CTGTTCGTGT AACTGGTCTT GCAGACAAAC AGTCTGATTT 4200

AACAGAAGAT TTCAAGGGTC CAGCAAAGAA AATTGCCTTA GATAGTGATG GAAACTTCAC 4260

CAAAGCAGCT CAAGGATTTG TCCGTGGGAA AGGTTTGACT GTTGAAGATA TCGAATTCCG 4320

TGAAATCAAG GGTGAAGAAT ATGTCTATGT CACTAAGGAA GAAATTGGTC AAGCAGTTGA 4380

AGCCATTGTT CCAGGCATTG TGGATGTCTT GAAGTCACTG ACTTTCCCTG TCAGCATGCA 4440

CTGGGCGGGA AATAGCTTTG AATACATCCG CCCTGTTCAC ACTTTAACTG TTCTCTTGGA 4500

TGAGCAAGAG TTTGACTTGG ATTTCCTTGA TATCAAGGGA AGTCGTGTGA GTCGTGGCCA 4560

TCGTTTTTTG GGACAAGAAA CCAAGATTCA GTCAGCATTG AGCTATGAAG AAGACCTTCG 4620

TAAGCAGTTT GTAATCGCAG ATCCATGTGA ACGTGAGCAA ATGATTGTTG ACCAAATCAA 4680

GGAAATTGAG GCAAAACATG GTGTACGTAT CGAAATTGAT GCGGATTTGC TGAATGAAGT 4740

CTTGAATTTG GTTGAATACC CAACTGCCTT CATGGGAAGT TTTGATGCTA AATACCTTGA 4800

AGTTCCAGAA GAAGTCTTGG TGACTTCTAT GAAGGAACAC CAGCGTTACT TTGTTGTTCG 4860

TGATCAAGAT GGAAAACTCT TGCCAAACTT CATTTCTGTT CGTAACGGAA ACGCAGAGCG 4920

TTTGAAAAAT GTCATCAAAG GAAATGAAAA AGTCTTGGTA GCCCGCTTGG AAGACGGAGA 4980

ATTCTTCTGG CGTGAAGACC AAAAATTGGT GATTTCAGAT CTTGTTGAAA AATTAAACAA 5040

TGTCACCTTC CATGAGAAGA TTGGTTCTCT TCGTGAACAC ATGATTCGTA CGGGTCAAAT 5100

CACTGTACTT TTGGCAGAAA AAGCTAGTTT GTCAGTGGAT GAAACAGTTG ACCTTGCTCG 5160

TGCAGCAGCC ATTTACAAGT TTGACTTGTT GACAGGTATG GTTGGTGAAT TTGACGAACT 5220

CCAAGGAATT ATGGGTGAAA AATACACCCT TCTTGCTGGT GAAACTCCAG CGGTGGCAGC 5280

TGCTATTCGT GAACACTACA TGCCTACATC AGCTGAAGGA GAACTTCCAG AGAGCAAGGT 5340

CGGCGCAGTT CTAGCCATTG CAGACAAATT GGATACGATT TTGAGTTTCT TCTCAGTAGG 5400

ATTGATTCCA TCAGGTTCTA ATGACCCTTA TGCCCTTCGT CGTGCAACTC AAGGTGTGGT 5460

TCGTATCTTG GATGCCTTTG GTTGGCACAT TGCTATGGAT GAGCTGATTG ATAGCCTTTA 5520

TGCATTGAAA TTTGACAGTT TGACTTATGA AAATAAAGCA GAGGTTATGG ACTTTATCAA 5580

GGCTCGTGTT GATAAGATGA TGGGCTCTAC TCCAAAAGAT ATCAAGGAAG CAGTTCTTGC 5640

AGGTTCAAAC TTTGTTGTGG CAGATATGTT GGAAGCAGCA AGTGCTCTCG TAGAAGTAAG 5700

CAAGGAAGAA GATTTTAAAC CATCTGTTGA ATCACTTTCT CGTGCCTTTA ACCTGGCCGA 5760

GAAGGCAGAA GGGGTTGCTA CGGTTGATTC AGCACTATTT GAGAATGACC AAGAAAAAGC 5820

TTTGGCAGAA GCAGTAGAAA CACTCATTTT ATCAGGACCT GCAAGTCAGC AATTGAAACA 5880

ACTTTTTGCG CTTAGCCCAG TCATTGATGC TTTCTTTGAA AATACTATGG TAATGGCTGA 5940

AGATCAGGCT GTCCGTCAAA ATCGTTTGGC AATCTTGTCA CAACTAACCA AGAAAGCAGC 6000

TAAGTTTGCT TGTTTTAACC AAATTAACAC TAAATAAAAT TTGATAAACG GACTTTATCT 6060

TATTACAAAG GAGAAGAAAT GGATCCGAAA AAAATTGCTC GTATCAATGA GCTTGCTAAA 6120

AAGAAAAAAA CAGAAGGCTT AACACCAGAA GAAAAAGTGG AACAAGCCAA ACTACGTGAG 6180

GAGTACATCG AAGGTTATCG CCGCGCTGTT CGTCACCACA TTGAAGGAAT CAAAATTGTG 6240

GACGAAGAAG GAAACGATGT TACACCAGAA AAACTACGCC AAGTACAACG TGAAAAAGGA 6300

TTACATGGCC GTAGTCTTGA TGATCCAAAT TCATAATAAT ACTCTTCGAA AATCAAATTC 6360

AAACCACGTC AGCTTCACCT TGCCGTACTT AAGTACAGCC TGCGGCTAGC TTCCTAGTTT 6420

GCTCTTTGAT TTTCATTGAG TATATGTATT CTTTCTTTTA ACAAAGATAG ATGAAACGAT 6480

AACAAAGAGA CTAGCAGTTT GTGTTTGCTA GTCTTTTTTC GCTAAAAAAG GAACCATAAT 6540

GGTTCCTAAA AACTATCATT AGTAACTTGC ACCGGCTGTA GCGTCTGCGT CACCACCGTG 6600

GCCTCCAGCA TCCCCTGAAT CAGAAGCGCC AGAAGTAGCA TCGGCGTCTC CATGACCTCC 6660

GGCAGCAGGA GCAAATGGTC CGCTACCACC CACCAAACGT TGACCAGTCT CTTTTAGGTA 6720

CCAGTCAAGC CATGGTTGGA AGTTAAAGAC GATTTCATTG ATACCAGCGT ATGATCCATC 6780

AGGATAGTAC ATTGCTTGGT AGTTGTGAGT GTTGATAACA CCTGCAGGAG AACCTGGAAC 6840

GATCGTACGG ACGTATTCTT GGTTTCCGTT GCGAAGTGTT CCGATAACCC ACTCTACGTT 6900

CTTCATACGT GCTGGTGGAA GAGAACCATG AACAGTCGAC ATACGGCTAC CTGATTGAGG 6960

TGGTACACGT TTAGCGAACA TAGTGTCTGG ATCTTGGTGA GCGTTGTTGT AGTAGAGGAA 7020

TTGGTTGTTG TCGTCAGCGT ATGTCAATTC AAATGGCATA GCTTTCAAGA ACATATCAAT 7080

TTGGTTAACT GTTAGGATAC CGTGGTCCAA TTTGACATAG GTATCACCAG AAACAGCACC 7140

AGTGAATGCT GCAACTTTTT CTACCCATTC TGGATCGTCA GGGTCAACTT CTGTGATGGT 7200

TGTAGCGATT GGTTTTCCAC AATCCAAGTC TTCTGATTCG ATTGGTTTTG GTTTTTTCAA 7260

TTTCGAAACG ACTCCTACGT ATTTAACAAA GTTATCTAAG CAAGTTTCAA GGAATTTAAC 7320

AGTGCCTTCG TTGGTGATAT TTCCGTTGTT ATCAAAAGCT TCCTTAGCTT TACCAAGAAG 7380

GAATTCGTTA CCTGGAAGCG TGTAGGCATT AACACCTGGA GCATCAAGGA TTTTACGAAG 7440

GTGAACTTGA GCACGTGATG TTCCTTGGTC ATAGTATGAT GCACCCACAA TCATAACAGG 7500

CTTGTTTTCA AATGGATGAA CTTCGTATGA AAGCCATTCA AGTACAGATT TGAGTGAAGC 7560

TGAGATAGTG TGGTTATGCT CAGGAGTAGC AATGATAACA CCATCTGCAC GAGTAATTTT 7620

GTTATATAAA TAACGTAATT GGAAACTTTC ATCCCATTTT TCATCTTGGT TAAACATTGG 7680

AACTTCGTCA ATTTCAAGAA CTTCTAATTC AAATTTGAGT TTGAAGTAGC GACGGATAAA 7740

TTCCAAGAGC TTACGGTTAT ATGATTGATC GTAGTTTGAT CCAACAAGTC CAACAAATTT 7800

CATTCTTTTT GGTCTCCTAT CTTACAAATT TTCCCAGTCA AAGTCTTCAG CATCTTTGCG 7860

AAGTAATTCT TGTGCATTAC GTAATTTTTC TGTGATTTTT ACAAAGATAC GGAAGTCATC 7920

AAAGATGGCA TCCAATTTCT TGATAACATC AAGGTCAACC AAGTCGCCAC TTGGGTTAAA 7980

TGCTTGAAGA GAGTGTGAGA GCAAGAATTC ATCTGGAAGA ACATTTGCCT TGATTTCAGG 8040

AGCATTCAAG ATTTGACGAA GTTGCAATTG GGCACGAGAT GAACCAAGCG TACCGTAAGA 8100

AGCACCTGTA ATCATGATTG GTTTGTTCAA AAGTGGGTAA ATACCATAAG ACAACCAAGC 8160

AAGAGCGCTC ATCAAAACAG CTGGAATAGA GTGATCATAC TCAGGAGTAC CGATAATAAC 8220

GCCATCTGCC TCTTCGATTT TAGCAGCAAT TTCCAATATT TCAGCAGGTA CTTGCTTGTC 8280

AGCTGGTTTG TTGAAGACAG GAATGGCCTT GATTTCAACA AGTTCAATTT CAGCTTTGTC 8340

AGTAAAGTGT TTTTGCATGT ATTGAAGCAA TTGACGGTTT GTAGAACGTT TTGAATTTGT 8400

TCCAACAATA GCAATAAGTT TTAACATGAG ATTTCCTTTC TCTTTTTACA TAATACAATT 8460

TTAAAATTCC ATTGAAACAG TTGTCTCTAT AGAGTAGGAA TTCCTGAAGA ACAGCTTAGG 8520

TGGCCTTCTT TATCGATGAG GATGACTTCG ATGCCCTCCA AACTTTCGAC TTGCCAGAGG 8580

ATAGAAGCAG GTCTTTCTCC AAAGAGTCGA GTCGTCCAGA TTTCGCCATC GACTGATTTA 8640

TCAGAGATGA TTGTTAGACT CGCTAGTTCC GTTTCAACAG GATATCCTGT TTGACTGTCA 8700

AAAATGTGAT GGTAATCTTG TCCATCGACG GTCAGGTGAC GTTCATAAAT GCCTGAAGTC 8760

ACGACAGATT TATTGACAAC AGGGATGGTC ATTAAATGAT TTCCCCTAGG ATTGGCTGGG 8820

TCTTGAATCC CGATTTGCCA TGGGTTATCC CCTCTTGCCT GATTTTTTCC AATGGTCAGG 8880

ATATTCCCTC CCAGATTGAT CAAGGCAGAA GTCACCCCCT CTTTCCTAAG AAATTGGGCA 8940

ACCTTATCCG CACTGTATCC TTTGGCTAAA CAACCTAGAT CGATCTTCAT TCCTTTCTGT 9000

TTTAAAAACA CAGTAGAAGT AGAAGAATCT AACTCGATAC CATGAGGATT GATTAGAGGC 9060

AGCACCGATT CAATTTCTTG AGGCTGGGCG ACCTTGGCAT CTGAAAAACC GATACGCCAG 9120

GTTTGAATTA AGGGACCAAT GCTGATATTG AGGTGGCTAG AGAGCGCTAG GCTATGCTCT 9180

AACCCAAGTG AAATCAGCTC AAACAGGTCT GGATGAACCG TGACGGGGGC TATTCCTGCT 9240

TGATAATTGA TTTCCATCAA CTCAGATTCT TGACTATTGG CGTTGAAGCG GTATTCAAGT 9300

TCTTTGAGCA AGTCAAAGGA TTTTTGGAGA AAGATATCGG CTTGCTCATC CACTAATGAA 9360

ATAGTGATAG TAGTCCCCAT TAGCCGTTCA GAATGTGAAC GAAGAGTCAA GCTACCAACT 9420

CCTTTCTCTT ATAGAAAATA AGTTGTAATA TCAAATAATC ATCTAAATTG AAGCCCTTAC 9480

ATTTCATTTT CATGTTATTA TAATACCATA AAGTTAGAAT TTTCACAAAC AAAATTTGGA 9540

AAAAGTCAAG AAATATGCTC ATAAAATTCA TCAGGCTTGA AAACAGGATA AATGGGGAAT 9600

TATTTTTGAT AAAAAATGCT GAAATAATAG TACCCCCCTT GTAAACGCTA ACGGTAAATG 9660

GTATACTAGT AAGGTAAATT TAGAATGAAG GCAGGAAATT TTTATGAGTA AAATCGTTGT 9720

AGTCGGTGCT AACCACGCTG GTACAGCATG TATCAATACC ATGTTGGATA ATTTTGGAAA 9780

TGAGAACGAA ATTGTTGTAT TTGACCAAAA CTCTAACATC TCTTTCCTAG GATGTGGAAT 9840

GGCTCTTTGG ATTGGTGAAC AAATTGACGG TGCTGAAGGC TTGTTCTATT CTGATAAAGA 9900

AAAATTGGAA GCTAAAGGTG CTAAAGTTTA CATGAACTCA CCTGTTCTTT CAATCGACTA 9960

TGATAACAAA GTAGTTACAG CGGAAGTTGA AGGAAAAGAG CACAAAGAAT CATACGAAAA 10020

ATTGATTTTC GCTACAGGCT CTACACCAAT CTTGCCACCA ATCGAAGGTG TTGAAATTGT 10080

TAAAGGAAAC CGCGAATTTA AAGCAACTCT TGAAAACGTA CAATTCGTGA AATTGTACCA 10140

AAATGCTGAA GAAGTTATCA ATAAACTTTC TGACAAGAGC CAACACCTCG ACCGTATCGC 10200

CGTTGTTGGT GGTGGTTACA TCGGTGTTGA ACTTGCTGAA GCCTTTGAAC GTCTTGGAAA 10260

AGAAGTTGTC CTTGTTGATA TCGTTGATAC TGTCTTGAAC GGTTACTATG ACAAAGACTT 10320

CACACAAATG ATGGCGAAGA ACTTGGAAGA TCACAACATC CGCTTGGCTC TAGGTCAAAC 10380

TGTTAAAGCA ATCGAAGGTG ACGGTAAAGT TGAACGCTTG ATTACTGACA AAGAAAGCTT 10440

TGACGTGGAT ATGGTTATCC TTGCAGTTGG TTTCCGTCCA AACACAGCCC TTGCAGGTGG 10500

TAAGATCGAA CTCTTCCGCA ACGGTGCCTT CCTTGTAGAC AAGAAACAAG AAACATCTAT 10560

CCCAGACGTT TACGCTGTTG GTGACTGTGC GACTGTTTAT GACAATGCTC GTAAAGATAC 10620

AAGCTATATC GCTCTTGCTT CAAATGCTGT GCGCACTGGT AACGTTGGT 10669

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7542 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

CGCGCTAATA GATACTTTAT GATAGAATAA AGAACAAGAT TGACAAGTAA GAGGAAACAT 60

TATGCAAAAT CAAACACTCA TGCAATACTT TGAATGGTAT CTGCCCCACG ACGGTCAACA 120

CTGGACGCGT CTGGCTGAAA ATGCTCCACA CCTAGCTCAT CTGGGGATCA GTCACGTCTG 180

GATGCCACCA GCCTTCAAGG CAACCAACGA AAAAGATGTC GGCTATGGGG TCTATGACTT 240

ATTTGACTTA GGAGAGTTCA ACCAAAAAGG GACTGTCCGC ACCAAGTATG GTTTCAAAGA 300

AGACTATCTT CAAGCCATTC AAGCCCTTAA AGCACAGGGA ATTCAACCTA TGGCCGATGT 360

AGTTCTCAAC CACAAGGCTG CTGCCGATCA CAGGGAAGCC TTTCAGGTTA TCGAAGTTGA 420

TCCTGTAGAC CGTACAGTTG AACTTGGAGA ACCCTTCACC ATCAATGGCT GGACTAGTTT 480

TACCTTCGAT GGTCGCCAAG ATACCTATAA TGGCTTCCAC TGGCATTGGT ACCACTTCAC 540

CGGTACAGAC TACGATGCCA AACGCAGTAA ATCTGGGATT TATCTGATCC AAGGGGACAA 600

CAAGGGCTGG GCCAACGAGG AATTGGTCGA TAACGAAAAC GGAAACTACG ACTACCTCAT 660

GTATGCCGAC CTAGACTTTA AACATCCTGA AGTCATCCAA AACATCTATG ACTGGGCTGA 720

TTGGTTCATG GAAACGACTG GTGTAGCTGG TTTCCGTTTG GATGCCGTTA AGCATATTGA 780

CTCTTTCTTT ATGCGCAACT TCATCCGCGA TATGAAGGAA AAATACGGTG ACGATTTCTA 840

TGTTTTTGGT GAATTTTGGA ACCCAGACAA GGAAGCCAAT CTGGACTATC TCGAAAAAAC 900

GGAAGAACAC TTTGACCTTG TCGATGTTCG TCTCCACCAG AATCTCTTTG AAGCCAGTCA 960

AGCTGGCGCA AACTATGACC TTCGTGGCAT TTTCACAGAT AGCCTGGTTG AACTCAAGCC 1020

TGACAAGGCT GTGACTTTTG TCGACAACCA CGATACCCAA CGAGGACAAG CCCTTGAGTC 1080

TACCGTTGAA GAATGGTTCA AGCCAGCAGC CTATGCCCTC ATTTTGTTAC GCCAAGACGG 1140

CCTTCCATGT GTCTTTTACG GAGACTACTA TGGGATTTCA GGGCAGTATG CTCAAGAAGA 1200

TTTCAAAGAA ATCCTTGACC GCCTCCTAGC CATCCGAAAA GATTTGGCCT ATGGAGAACA 1260

AAATGACTAC TTTGACCATG CTAACTGTAT CGGTTGGGTA CGTTCAGGTG CTGAAAATCA 1320

ATCCCCAATC GCAGTCCTTA TCTCAAATGA CCAAGAAAAC AGCAAGTCAA TGTTTGTCGG 1380

TCAAGAATGG ACTAATCAAA CCTTTGTAGA TTTACTTGGT AACCACCAAG GTCAAGTTAC 1440

AATTGATGAG GAAGGTTATG GACAATTCCC TGTCTCAGCT AGATCCGTAA GTGTCTGGGC 1500

AGTCAATACC ATCTAATAGC TCATAATAAC CAAGCTAGGT CCAAGCGGAT TTGGCTTTTT 1560

TGTATTCACA AAAAGACCTA CCCAAATGGA TAGATCTTTA CTTGATTACA ATTTACCTGC 1620

TACTGCATCC AACAATTCTT GGATCTTAGG TTGGTTGCTT CCTCCTGCCA TGGCCATATC 1680

TGGTTTACCA CCACCACGTC CATCGATGAT TGGTGCTAAT TCTTTGACAA GGTTTCCTGC 1740

ATGAAGGTCT TTTGTCTTGC TTGCTACAAG GACATTGACT TTGTCACCGA TAGCGGCAAC 1800

TAGGACAAGA AGATCAGAGT AGTCTTTTTG TTTCCAGTTA TCTGCAAAAG TACGAAGGGC 1860

ACCGGCATCG GATACAGACA CTTGACTAGC AATGTAACGA TGACCGTTGA CTTCCTTAAC 1920

ATCTTTGAAG ATATCGCCTG CGGCTGCAGC TGCGGCTTTT TCTTTCAACT CAGCATTTTC 1980

TTTTTGAAGT TGACGAAGTT GTTCTTGAAG TCCTTCTACC TTGTGAGGTA CTTCCTTGAC 2040

TTGAGGTGCT TTCAAGGTTG CTGCGATAGC TTTAAGAGCA TCCTCTTGTT CACGATAGGC 2100

TTCAAAGGCT TCCTTACCAG TCACTGCCAA GATACGGCGA GTTCCTGAAC CGATTCCTTC 2160

TTCTTTGACA ATTTTGAAGA GACCAATCTC AGAAGTGTTG TCAACATGAG TACCACCACA 2220

AAGTTCAATA GAGTAGTCAC CGATAGTCAC GACACGAACT TCCTTGCCGT ATTTCTCACC 2280

AAAGAGGGCC ATAGCTCCCA TTTCTTTAGC AGTGTCAATA TCCGTTTCAA CTGTCTTCAC 2340

TTCAAGTGCT TCCCAAATTT TCTCGTTAAC TTGCTGTTCA ATCGCACGAA GTTCCTCAGC 2400

AGTTACTGCT TGGAAGTGGG TAAAGTCAAA GCGAAGGAAT TCAACTTCGT TAAGAGATCC 2460

TGCCTGTGTT GCGTGGTTTC CAAGGATATT GTGAAGGGCA GCGTGAAGCA AATGAGTCGC 2520

AGTGTGGTTT TTCATGACAC GGTGACGGCG ATTGCTATCA ATTGCCAAGG TATATTCTTG 2580

GTTCAAGGCA AGCGGTGCAA GGACTTCAAC TGTATGAAGG GCTTGACCAT TTGGGGCTTT 2640

CTGAACATTG GTCACAGTAG CCACAACCTT ACCTGACTCA TCCAAGATTT GTCCGTAGTC 2700

AGCTACCTGT CCACCCATTT CAGCATAAAA TGACGTTTCC GCAAAGATAA GAGAGGCAGT 2760

TCCTTCTGAA ACAGCTCCTA CTTCTGCATT GTCAGCAACG ATAGCTACCA ATTTAGAAGA 2820

CAATTGGCTA GCATTGTAGT TGAAGACACT TTCTACAGTG ATGTTTTGAA GAGTTTCATT 2880

TTGCATACCC ATTGAGCCAC CCTTGACAGC TGACGCACGC GCGCGTTCTT GCTGTTCTTT 2940

CATGGCTGCT TCAAAACCTT CACGGTCTAC AGTCATACCA GCTTCTTCAG CGATTTCTTC 3000

AGTCAATTCA ACTGGGAACC CATAAGTATC ATAGAGTTTG AAGACATCTG AACCAGCGAT 3060

AACAGATTGA CCTTTTTCTT TCAAGTCTGC TACAATGCCT TGGGCAAAGT GTTGACCTGA 3120

GTGAAGGGTA CGGGCAAATG ATTCTTCTTC GCTCTTAACG ATTTTCTCAA TAAAGTCACG 3180

TTTCTCAAGC ACTTCTGGGT AGTAGCTTTC CATGATTTTT CCAACAGTTG GAACCAATTT 3240

GTAAAGGAAA GGCTCGTTGA TACCCAATTT TTGACCATGC ATAGAAGCAC GACGGAGAAG 3300

ACGACGAAGA ACATAACCAC GACCTTCATT TCCTGGAAGG GCACCATCAC CGATAGCAAA 3360

TGAAAGAGAA CGAATGTGGT CTGCGATAAC CTTGAAGCTC ATGTTGTCGC CATCTTGGTC 3420

ATAAACCTTA CCAGACAATT TCTCGACTTC ACGGATAATC GGCATGAAGA GGTCCGTTTC 3480

AAAGTTGGTC TTAGCCCCTT GGATAACGGC CACCAAACGC TCCAAACCAG CGCCCGTATC 3540

AATGTTCTTA TGTGGCAATT CCTTGTATTC GCTACGAGGA ACAGCAGGGT CTGCGTTAAA 3600

TTGTGACAAA ACGATGTTCC AGATTTCAAT ATAACGGTCG TTTTCAATAT CTTCTGCAAG 3660

CAGGCGAAGA CCGATATTTT CTGGGTCAAA GGCTTCCCCA CGGTCAAAGA AGATTTCTGT 3720

ATCTGGTCCA GAAGGTCCCG CACCGATTTC CCAGAAGTTG TCCTCAATTG GAATCAAGTG 3780

ACTTGGATCC ACTCCCACTT CAATCCAGCG GTTGTAAGAA TCTTTATCGT CTGGATAGTA 3840

GGTCATGTAA AGTTTTTCAG CAGGGAAATC AAACCATTCA GGGCTTGTCA AAAGCTCATA 3900

AGCCCAAGTG ATAGCTTCGT CACGGAAGTA ATCCCCGATA GAGAAGTTCC CCAGCATTTC 3960

AAACATGGTA TGGTGACGCG CGGTCTTCCC TACGTTTTCG ATGTCGTTGG TACGGATAGC 4020

CTTTTGGGCA TTGGTAATAC GTGGATTTTC AGGGATAATG GTCCCGTCAA AGTATTTCTT 4080

AAGGGTTGCT ACCCCAGAGT TGATCCACAA AAGAGTTGGG TCATTTACAG GAACCAAACT 4140

TACTGATGGT TCTACTGAGT GACCTTTGGT CGCCCAGAAA TCAAGCCACA TTTGGCGTAC 4200

TTGTGCACTA GATAGTTGTT TCATATTGTC TCCTTATTCA CTTGTTTAAT GTGATTGGCT 4260

TTCCAGCATT TCCACATAGT CAATCGCGAC ACAGAGGGAA ATGACTAGGT CTGCATAAGC 4320

GTCTTCAAGA ACCGTTACGG TATAGGTAGA AGTCAGATGG AAGAGTTCCT TCTTAATTTC 4380

CGCAATCAAC TGATCGCGAT CATCCAGCAA TTTGAAATTC AAATCCCAGA TATTGCCCTC 4440

GATACGAAGA CCTAGATTAT CAAACTCATA CTTATCTCGC CAGAAGGTCA ACTTCTTACG 4500

AATGACAAAA CTCGAGCCAT CCCGAAGCTG AATTTCAAAA CGAGGAAGCA AGGTCAAGAT 4560

TTCTTTACTA ATCTCACTGA CTTGTTCACC AGCCGCATCA TAGATGGTAA AGGTTTTAGG 4620

AATCTTAAAA AATGATCCCT CCACCTGATA GGCAATTTCT CCCCTGTCAT CCTTGATAGC 4680

GAAGCGTTCG CCTCCAAGAC GAAACTTTTG TTTGACAAGA AATGTTTTCA TCAACACCTC 4740

CAAAAATCAA AAGACAAGCT CATATCACGA AGGGCGAAAA ACCGCGGTAC CACCTTCATT 4800

CAATGAACTT GTCATTCTCT TGTTCTTATG CAATTGTATG ATTGAGTAGC ATGACTTCCT 4860

AGCTTAGATG GCTCGCAGCA CCGCCATTTC TCTGGACTAA GACAAGTGAA AATCAATTCT 4920

CAACTTTCTT ATTATAACGT TTTTTTAAGC TTGCGTCAAC TGGAAATGAT CTCCGTTGAA 4980

TTAGACCAAT TCCCTACATC TCTGATTACT TTTTCAGGAT ATATTTTTTC TTACTGCCAT 5040

TTTTCTTTTT ATCCCAAATT TTCATATTAC TAAACACAGC TACTAGAATA TTTCCAAATA 5100

TAAAGGTGCC TATCACCCAA TATATGGACT CAGTTGTTAG GTATTGTCGA TCCAAGCCAT 5160

CCTTTAAATG GAATAGTATA GCAGTTTGGT TAACAATCAT AAAGGTTGGC CAGAAACTTT 5220

TTTTGAAAAA AGTAGACATT TTCATTATTT GTTGCCGCTT TCTGTAAGGT TAATACTCAA 5280

TAAAAATCAA AAAGCAAACT AGGAAGCTAG CCTCAAGCTG TACTTGAGTA CGGCAAGGCA 5340

ACGCTGACGT GGTTTGAAGA GTATAGGCTT AGTATACTAC TAGGCAAGCA AATAAACAAA 5400

TAAACAACTA GAATAGAAAA AGATAGGGCT CTAAAAACTG ACTTCTATTC CTTAAAAACG 5460

AACCAGCTTG ACTGATTCGT CTTCTTACGT TTATCTCCTA CTTCCGATAC ATTTTAAACT 5520

GTAGGAAGAG GTCGCTATAT TTCCCTGTCC ATTTATGGTC AAATTTCTCA TAAACTTCTA 5580

GGTGTTTCAT GGTTTCAACA TCGGGATAGA AGGCCTTATC TTCCTTTGTT TCCTCTGGGA 5640

GCAATTCCTT CGCTGGTAGG TTTGGTGTTG AATAGCCGAC ATACTCCGCA TTTTGGAGAG 5700

CATTTTCAGG TTTCAACATA AAGTTGATAA AGGCATAGGC TGAGTTTTGG TTTTTAACTG 5760

TTTTGGGAAT GACCATATTG TCAAACCAAA GATTGCTGGC CTCTGTCGGT ACCACATAAC 5820

GTAGATTTTC ATTTTTTTCT AACATTTGGC TGGCTTCACC AGAGAAGGTC ACGCCGATTG 5880

CAACATTATT CTGAATCATA TAGCCCTTCA TCTCGTCCGC AACGATAGCC TTGATATTTG 5940

GAGTCAGTTT GTAGAGCTTA TCCACTGTCT CTTCCAACTG CTGCAGATCC TTGGAGTTGA 6000

GGCTGTAGCC GAGGGAATTG AGTCCTAGTC CCAGCACCTC ACGCGCCCCA TCAAAGAGCA 6060

TGATAGAATT CTTATACTCC GGCTTCCAAA GGTCATCCCA ATGCTCAGGC GCTTCATCTA 6120

CCATGGTTTC GTTGTAGACA ATTCCTAAGG TTCCCCAGAA GTAAGGGATG GAGAATTTAT 6180

TACCTGGGTC AAAGGACTGG TTGAGAAACT CTGGTCCGAT ATTTTCGATT CCTTCAATTT 6240

TTGAATAATC AAGCGGAACC AAGAGGTCTT CGTCCTTCAT CTTGTTAATC ATGTATTCAC 6300

TTGGAATGGC AATATCGTAG GTCGTTCCAC CCTGCTTTAT CTTAGTGTAC ATGGCTTCGT 6360

TGGAGTCAAA AGTCTCGTAC TGAACTTGAA TTCCTGTTTC TTCTGTAAAC TGAGTCAAGA 6420

GTTCAGGATC GATATAGTCT CCCCAGTTAT AGATAACCAA TTTTTGACTA TCTCGACTAT 6480

TGATTTTACT ATCTAAATGA GTCGCAATTC CCCACAAGAC AAGGATAATC GCTGCAATTC 6540

CTGCTAAAAA TGAATAGATT TTTTTCATGC TTGCTCCTCC TTCTCACGAG AGATAAAGTA 6600

ATAACCTACA ACTAGGATAA TACTAAAGAG AAAGACTAGA GCAGACAGGG CATTGATTTC 6660

TAAGGAAATC CCCTTGCGAG CACGAGAGTA AATCTCGACT GATAGGGTTG AAAAGCCATT 6720

TCCTGTTACA AAGAAGGTCA CGGCAAAGTC ATCTAACGAA TAGGTGAAGG CCATGAAATA 6780

ACCAGTAATG ATAGACGGAG TCAGGTAAGG AAGCATGATT TCCTTGAACA TCTGAAATTG 6840

ACTAGCTCCC AAGTCATAGG CCGCATGAAT CATGTCGCCA TTCATTTCCT TGAGTCGAGG 6900

CAAGACCATC AAGACCACGA TAGGAATGGA GAAGGCCACG TGACTAGATA GAACGGTCAA 6960

AAAGCCAAGT GAAAACTTGA GTTGGGTAAA GAGAATCAAG AAGCTAGCAC CAATCATAAC 7020

GTCAGGCGCA ACCATGAGGA TATTATTGAG TGATAGAAAG GCTTCTTGGT ATTTCTTACG 7080

AGACTGGTAG ATGTAAATGG CACCAAAAGT CCCGATAATG GTCGCTATCA AGGCTGATAG 7140

GAAGGCCAAG AAAAATGTCT GAGCCAAAAT CAGCATGAGT CTCCCATCTC CAAACATGGT 7200

TTCAAAGTGA GTCCAGCTAA AACCTGTAAA GCTATTCATA TCATCACCAG CATTAAAGGC 7260

ATAGCCAATC AAGTAAAAGA TAGGCAGGTA GAGGACCAGA AAGACCAGTC CCAGATAAAG 7320

GTTGGCAAAT TTTTTCATCG TTCTCTCCTT TCCTTAGTCA CCCACATGGT GATGAACATG 7380

GTCAGGATGA GAATCACACC GATGGTTGAA CCCATACCAT AGTTGTCATT GGTTAGAAAA 7440

TTCTGCTCAA TAGCCGTCCC CAAGGTGATA ACGCGTTCCC ACCAATCAAA CGGGTCAGCA 7500

TGAAGAGACT CAAACTTGGG ATAAAGACCG ACTGAACCCC GG 7542

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9223 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

AAAACCAAAT TCCGGTATTT TAACCTATGC TGTAAATACC ATGAAGTCTG TCATGACAGA 60

TCAGGTCTAT AACATTAAGG TTGAGACAGA AAATGGAAAT TATGTTGGTG AAGCTAGCCA 120

TGTTTTGGTC CTTTTGACAA ATTACTTCGC TGATAAGAAA ATCTTTGAAG AAAACAAGGA 180

CGGCTATGCC AACATTTTGA TTCTGAAAGA TGCCTCTATA TTCTCCAAAT TATCCGTCAT 240

TCCTGATTTA TTAAAAGGGG ATGTTGTCGC AAATGATAAT ATCGAGTATA TCAAAGCGCG 300

TAATATTAAA ATCTCTTCAG ATAGTGAATT GGAGTCAGAT GTTGACGGAG ATAAATCAGA 360

TAACCTACCT GTAGAAATCA AAGTCCTAGC TCAGCGAGTA GAAGTATTTT CAAAACCGAA 420

AGAGGATTAG TATATAGAGA AAGCCTTTTT TAAGGCTTTT TGTATACTTT AAAAGATAGT 480

TCCTTTAACA ACGGACATTC CTTGCAAATA GTTTTACAAA AATAGTATAC TGGATTCATT 540

GAGTTTGAAA ACGTTTGCGT AAAATTTGAA TGAATACTTT AGGAGACAAA TTGATGGAAT 600

TGAGTGCTAT TTACCATAGG CCTGAGTCGG AGTATGACTA TCTTTATAAG GATAAGAAAC 660

TCCATATTCG AATTCGAACT AAGAAAGGGG ACATTGAAAG CATCAACTTG CACTATGGGG 720

ACCCTTTTAT CTTTATGGAG GAGTTTTATC AGGATACAAA AGAAATGGTC AAGATAACTT 780

CTGGTACCTT ATTTGACCAT TGGCAGGTTG AAGTGTCAGT TGACTTTGCA CGTATCCAGT 840

ATCTCTTTGA GCTCAGAGAT ACAGAAGGTC AAAATATTTT GTATGGCGAT AAAGGGTGTG 900

TGGAAAATTC TCTAGAAAAT CTTCATGCAA TTGGGAATGG ATTTAAGTTG CCTTAGCTTC 960

ATGAGATTGA TGCCTGCAAG gTTCCTGACT GGGTTTCAAA TACGGTATGG TATCAGATAT 1020

TTCCTGAAAG ATTTGCCAAT GGCAATGCTC TATTAAACCC AGAAGGGACT TTAGACTGGG 1080

ATTCATCTGT CACACCTAAG AGCGATGATT TCTTTGGTGG TGATTTACAG GGGATTATTG 1140

ATCATATGAA TTACTTGCAA GACTTGGGTA TTACTGGACT ATATCTTTGT CCCATCTTTG 1200

AATCTACAAG CAATCACAAG TACAATACGA CAGATTACTT TGAAATTGAC CGTCATTTTG 1260

GAGACAAGGA GACCTTTCGG GAACTGGTGG ATCAAGCGCA TCATCGTGGC ATGAAAGTCA 1320

TGCTGGATGC GGTATTTAAT CATATTGGTT CGCAATCTCT TCAATGGAAA AATGTCGTCA 1380

AAAATGGTGA ACAGTCTGCT TATAAGGATT GGTTCCATAT TCAACAATTC CCAGTGACAA 1440

CTGAAAAGCT AGTTAATAAG AGAGACTTAC CCTATCATGT TTTTGGTTTC GAGGACTATA 1500

TGCCTAAGCT AAATACAGCC AATCCAGAGG TCAAGAATTA TCTTTTAAAG GTTGCGACTT 1560

ATTGGATTGA AGAGTTTAAT ATCGATGCTT GGCGTTTGGA TGTGGCTAAT GAGATTGACC 1620

ATCAGTTCTG GAAGGATTTT CGTAAGGCAG TTTTAGCTAA AAATCCTGAT CTTTATATCC 1680

TAGGAGAAGT CTGGCATACA TCTCAGCCTT GGCTAAATGG AGATGAGTTC CATGCCGTCA 1740

TGAATTATCC TTTATCTGAT AGTATCAAGG ACTATTTCTT ACGAGGAATT AAGAAGACAG 1800

ACCAGTTCAT CGATGAAATC AATGGAGAGT CTATGTATTA CAAGCAGCAG ATTTCAGAGG 1860

TCATGTTTAA TCTCTTGGAT TCACATGATA CAGAGCGAAT CCTGTGGACG GCCAATGAAG 1920

ATGTTCAACT GGTTAAATCA GCCTTAGCCT TTCTCTTTTT ACAAAAAGGA ACACCGTGCA 1980

TTTATTACGG AACCGAGCTA GCCTTGACTG GAGGACCAGA TCCAGATTGT CGTCGTTGTA 2040

TGCCTTGGGA ACGTGTATCA AGTGACAATG ATATGCTGAA CTTTATGAAG AGGCTGATTA 2100

AAATTCGGAA ATACGCGTCA GTAATCATTT CGCATGGCAA GTATAGCCTT CAAGAAATCA 2160

ACTCTGATCT AGTAGCTCTG GAATGGAAAT ACGAAGGACG GATCCTCAAA GCAATATTCA 2220

ACCAATCAAC AGAAGATTAT CTTTTAGAGA AAGAAGCAGT AGCACTAGCA AGCAATTGCC 2280

AAGAATTGGA TAATCAGCTT GTCATCTCTC CAGATGGATT TATGATTTTC TAAAAACTAG 2340

TTGATGAAGA TTATGGTACA TTTCATACCT TATATAGTAT AATAAGGCTA GTTACTAAAC 2400

TTGTAAAGGA GAACTTAAAT GAATTGTAGA GGACATGAAA CAAGACAAAG AATTGTTAGA 2460

GATTTTGAAG TTCAGCCTAA AGCACATATT AAGCTGTTAG CAAATCAACA AAAACATAGT 2520

GATGCAGGAG CAACTATTGA AGATGAATAT TATGTATTTA TCGCTGAGAG TAAAATTGAT 2580

GGCAAGAAGG AAGTTATTCA GTGTTGCATG GGTGCGGCAA GGGATTTTTT AGAACTAATT 2640

AATCACAAAG GGCTACCTCT TTTTAATCCG CTTGTAGGTG ATTCTCATGT AAATAATAGA 2700

CAAGAATATG ACAATACAGG GAGTGGAAAT TTATAACCTG AAAAGTGGAA TGAAACTGCA 2760

AAGCAGCTTT ATAATGCTAT AATGTGGTTG ATTATTTTAT GGAATGCTAA GCCGGATACA 2820

CCTTTATTTA ATTTTAAAGA CGAAGTAATT AAGTATAAAA CATATGAGCC TTTTGAAAGC 2880

AGTATAAAAA GAGTAAATAC TACTATAAAG AATGGTAGTA AAGGGAAAAC TCTGACTGAG 2940

ATGATTAATG GCTACAGAGC GGATAACGAT ATTAGAGATG AAATTTGTAA CTTTAATATT 3000

CTGAAAAATA AAATTCGTGA TATGAAAAAC CAACAAGGAA ATACAATGGA ATCTTACTTT 3060

TAGTTATTGT TGAATTTTGG GTATTCTATA AAATATCCTA ATTGAGATTT AAATAGTAGA 3120

CTATACAATA TAGTTAAAAT ATCAGTAAAA ACAACACTTT ATTGAGGTAT TGGATACGCT 3180

TTGCTAATAG CCTAATAATC ACATGTGGAG TGTTGCTACA ACGAAAAAGG TGATAATCCT 3240

TGATTTCAAG CTATTTTATA AGCATTTTGT CTTTGTAGAT AAAGGCAATT TTGACAATAA 3300

AAATCCTAAA AGGTGAATCG TTATAGATGT ATTTGTAGAT ATCGTTTGCG CATCGAAAAA 3360

ATTAATACAA GAATAAATAT TTATAGCTCT TTAGGTGACT TTTATAGAAG TAAAGTTTAG 3420

GATAGAAAAA CAAGAAATAA CGCACCATTT TTGGTGCGTT ATGCTTTTTT ATGCTATAAT 3480

GCATTTATAA AAATAAAGGA GTTTGCTATG ATTGGAAAGA ACATAAAATC CTTGCGTAAA 3540

ACACATGACT TAACACAACT CGAATTTGCA CGGATTGTAG GTATTTCACG AAATAGTCTG 3600

AGTCGTTATG AAAATGGAAC GAGTTCAGTC TCTACCGAAT TAATAGACAT CATTTGTCAG 3660

AAGTTTAATG TATCTTATGT CGATATTGTA GGAGAAGATA AAATGCTCAA TCCTGTTGAA 3720

GATTATGAAT TGACTTTAAA AATTGAAATT GTGAAAGAAA GAGGTGCTAA TCTATTATCT 3780

CGACTCTATC GTTATCAAGA TAGTCAGGGA ATTAGCATTG ATGATGAGTC TAATCCTTGG 3840

ATTTTAATGA GTGATGATCT ATCTGATTTG ATTCATACGA ATATCTATCT AGTAGAAACT 3900

TTTGATGAAA TAGAGAGATA TAGTGGCTAT TTGGATGGAA TTGAACGTAT GTTAGAGATA 3960

TCTGAAAAAC GGATGGTGGC CTAATGGAAA TCCAAGATTA TACTGATAGT GAATTCAAAC 4020

ATGCTTTAGC AAGGAATCTT CGTTCACTGA CAAGAGGAAA AAAGTCCAGT AAGCAACCTA 4080

TAGCGATTTT GCTTGGAGGG CAAAGTGGTG CCGGTAAGAC TACAATTCAT CGTATTAAAC 4140

AGAAAGAATT TCAAGGAAAT ATTGTTATCA TAGATGGTGA TAGTTTTCGT TCTCAGCATC 4200

CACACTATTT AGAACTGCAG CAAGAATATG GCAAAGACAG TGTAGAATAT ACCAAAGATT 4260

TTGCAGGAAA AATGGTAGAG TCTTTAGTAA CAAAATTGAG TAGTTTGAGA TACAATCTTT 4320

TGATAGAGGG AACTTTACGA ACAGTTGATG TTCCAAAGAA AACAGCACAA CTCTTGAAAA 4380

ATAAGGGATA TGAAGTACAA TTGGCCTTAA TTGCGACAAA GCCTGAATTG TCGTATCTAA 4440

GTACTCTTAT CCGTTATGAA GAACTGTACA TTATCAATCC AAATCAAGCA CGCGCAACTC 4500

CAAAAGAACA TCATGATTTC ATTGTAAATC ATCTAGTTGA TAACACACGA AAATTGGAAG 4560

AACTAGCTAT CTTTGAAAGA ATTCAAATTT ACCAACGAGA TAGAAGTTGT GTATATGATT 4620

CAAAAGAAAA TACAACTTCA GCAGCAGATG TTCTTCAAGA GTTACTCTTT GGGGAGTGGA 4680

GTCAGGTAGA GAAGGAGATG TTGCAGGTGG GGGAAAAGAG ACTTAATGAA TTACTTGAAA 4740

AATAAACAAT TGATATTTTT AGGAGAATAG AAATGAGAGG GTTTAATAAC AAGATAAAGT 4800

CTGTTTATCA AGAACTAACA AATTCCAAAG AGAAATTCGG TAGCTTTCAC AAGACTTTAA 4860

TTCATTTGCA TACACCTGTT TCTTATGATT ACAAGCTATT TTCTAATTGG ACTGCAACGA 4920

AATATAGAAA AATTACTGAA GATGAACTAT ATGATATATT TTTTGAAAAT AAGAAAATAA 4980

AAGTTGATAA GACAATTTTT TTTAGTAATT TTGATAAGGT TGTTTTTTCT AGTTCAAAAG 5040

AATATATTAG TTTTCTTATG TTAGCAGAGG CAATCATAAA AAATGGAATA GAAATAGTTG 5100

TAGTAACTGA TCATAATACT ACCAAAGGTA TTAAAAAGTT ACAAATGGCA GTCTCAATCA 5160

TAATGAAAAA TTATCCGATT TATGATATAC ATCCTCATAT TTTACATGGA GTAGAAATTA 5220

GTGCAGCAGA TAAATTGCAT ATTGTATGTA TATATGATTA TGAACAAGAA TCATGGGTTA 5280

ATCAATGGTT AAGTGAAAAT ATTATAAGTG AGAAAGATGG AAGTTATCAA CATTCACTGA 5340

CTATAATGAA GGATTTCAAT AATCAAAAAA TAGTTAACTA TATTGCTCAT TTCAATAGTT 5400

ATGACATTTT GAAAAAAGGT TCTCACTTAT CAGGTGCATA TAAACGAAAA ATTTTTTCTA 5460

AAGAAAATAC ACGATTTTGG AGTTTAATAT TAACTCGAAA GAATCTTCGC AACAACTTGA 5520

TATTCTCTAT AAAGAAGTTG GTGTATTAAG TTTGGGACAA AAAGTTGTAG CCATGCTTGA 5580

TTTTTTATTA GCATATAGTG ATTATTCTAA AGACTTCAGA CCATTGATTA TTGATCAGCC 5640

TGAAGACAAT CTAGACAATC GTTATATTTA CAGGCATTTA GTTCAGCAGT TTAGAGATGT 5700

GAAAGCTCAA CGTCAAATTA TTTTAGCAAC ACATAATGCT ACAATTGTAA CAAATTCTAT 5760

GACAGATCAA GTTGTTATTA TGGAGTCAGA TGGAGTTAAC GGATGGATTG AATCACAGGG 5820

ATATGTTAGT GAAAAATATA TAAAAAATCA TATCATCAAT CAATTAGAGG GAGGAAAAGA 5880

TTCCTTCAAG CATAAAATGT CTATATATGA GACGGCTTTA TCAGAGTAGA GTCAGAAAAA 5940

GTAGGTTAGA AATTTAGCCT ACTTTTTTCT TTGTCCGACA GGCATAGTGT ACATCTGAGG 6000

TCCAAGTCCT CTGTGGATAT TTGCTGCAGA TGAAACCAAT AGCGACTCCT AAGCCTGAAT 6060

ATCGTGAGGT AGGGGGGATA GGAAGGAATT AGCGAAATCA AGGTTCTACA AACAGAATCG 6120

TGACTTGAAG CCATATATAG CGGATGAGGA ACTCTAAAAT CCAAATAGGT GTCGTAACCT 6180

ATATACGTAA ATTACGAGAG TAAACTAGGA AAGATGTACG GCTTATTCCG TGAGCGTTTA 6240

GGACGTAGTA CAACGAATCA TGGGAGTCAG CTGAACACAT AGTATTGAAG AAATTTCTGT 6300

AATGGAAATG GAGCGAAGAA GTGAACAATT AAATGAATAC CTCTCTAATT AAATTTGTCA 6360

ATTCTAATTC CTGGTATGAA AAGACAGTGA CCTGAAAATG TAAACGATGG GAGCTGATCA 6420

TAAATATAGG ACGGTACATG CAGTGGTGTT AGAGATTAGT CCTTACTTGA TTTGTGATAA 6480

CTTCCCCAAA TTTCTTCTGC TATACTTTTC TCAACTTTTA AAAATCCAAC TAAGAATTTT 6540

ACCTGGGGGT TTGGGGGCGG AGCACTAAGT TATCTTATCG TTAGCTGTCA AAACTGGTAG 6600

GTTTTGATAG GCTGGCGATA TGATTTTTGG GATATTGTGG ACACAATATC TGAGCTCGCA 6660

AAGCCTTACA AGAATGAAAA TCAGTTGTTG GAAAAGTGTA CTGACATTGT ATGGTAGCTC 6720

ACATTGTCAG TACAAGTATT TTGGAAAGGA AGTAGCAGTA TGAAACGAGA TGTGCGTGAT 6780

ATTCGGAAAC AATTTCGTTT AACAGAAGCA GAAGAAAAGC AAATTCTAGC TTTGATGAGA 6840

GAGCGGGGAG AGACTAATTT CTCTGATTTT CTTCGTAAAA GTTTACTTTC CTCTGATTTA 6900

CAAAAACAGA TGGAGACATG GTTTGCCCTC TGGCAATCCC AAAAACTAGA ACAAATCAGT 6960

CGTGACGTTC ATGAAGTTTT AATCTTGGCA CAGTCAGAAC GTCAAGTCAC CCAAGAGCAT 7020

GTATCTATTC TCTTAACGTG CGTGCAGGAA TTGATTCAAG AGGTTGCAAA CACCATACCC 7080

CTCAGTAAAG AATTTCGTGA GAAGTACATG AGGTAAGCAC ATGGAACATC GTTACCGAAC 7140

CAATCTCAAG AAAGTGTTTT TGTCTGATAG TGAGTTGAAC CAACTAAATA TAAATATCGA 7200

TCAAAGTGGT TGTAAATCCT TTTCTGAATA TGCGAGACGA ACTCTACTCG ATCCTGGTAT 7260

GAATTTTATC ACGATTGACA CAAACGGTTA CCAAGATTTA GTGTTTGAGT TAAAGAGGAT 7320

TGGCAATAAT ATCAACCAGA TTGCTCGAAG TGTTAATCAA TCTCAGTTAA TTTCTGGTGA 7380

AGAATTGCAG GAGTTGAAAA AAGGAATTGG TGAATTGATA AAAGAAGTTG ATAAGGAATT 7440

TAATCTGCAA GCGCAGAAGC TAAAGGAGTT CCATGGTCAT CACTAAACAC TTTGCCATTC 7500

ACGGAAAGAG TTACCGCAGA AAGCTTATCA AGTACATTCT CAATCCTGAG AAAACCAATA 7560

ATCTTGCCTT GGTGTCGGAC TATGGCATGA AGAATTTTCT GGACTTTCCT AGCTATGAGG 7620

AAATGGTGCA GATGTATCAT GAAAATTTCA TCAGCAACGA TACGCTTTAC GATTTTCGCC 7680

ACGACAGGAT GGAAGAAAAT CAACGAAAAA TACACGCTCA CCACATCATT CAGTCTTTCT 7740

CGCCAGAGGA TCATATCACT CCTGAACAAA TCAATCGGAT AGGTTATGAG ACTGTGAAGG 7800

AATTAACTGG TGGCAAATTT CGTTTTATCG TTGCGACCCA TGTTGATAAA GACCACCTGC 7860

ACAATCACAT CATTATCAAT TCAGTAGATA GCAATTCTGA CAAAAAGCTC AAGTGGGACT 7920

ACAAGGTGGA GCGAAATCTT CGCATGATTT CTGACCGTTT TTCTAAAATC GCAGGTGCTA 7980

AAATCATTGA GAACCGCTAT TCTCACCAGC GGTATGAAGT CTATCGTAAG ACTAATCACA 8040

AGTATGAACT CAAGCAGCGA CTCTATTTTT TGATGGAACA TTCTAGGGAC TTTGAGGATT 8100

TCAAAAAGAA TGCTCCGCTA CTACATGTGG AGATGGATTT CCGTCACAAG CATGCCACCT 8160

TTTTTATTAC GGACTCAACT ATGAAACAGG TGGTGCGTGG CAAGCAACTC AATCGCAAGC 8220

AGCCTTACAC AGAAGAATTT TTTAAGAACT ACTTTGCCAA AAGAGAAATA GAAAGTCTCA 8280

TGGAATTTTT ATTGCTGAAA GTTGAGAATA TGGATGATTT ACTTCAGAAA GCAAAACTTT 8340

TTGGACTAAC TATCAATCCT AAACAAAAGC ATGTTTCTTT TCAATTTGCA GGAGTGGAGG 8400

TAAAGGAGAC AGAGCTAGAC CAGAAAAATC TTTATGATGT AGAGTTTTTC CAAGATTATT 8460

TTAAAAATAG AAAAGATTGG CAAGCTCCAG AAACTGAGGA TTTCGTTCAA CTTTATCAAG 8520

AAGAAAAGTT ATCCAAAGAA AAAGAACTTC CAAGCGATGA GAAGTTCTGG GAGTCCTATC 8580

AAGAGTTCAA GAGTAACAGA GATGCCGTTC ATGAATTTGA GGTGGAGTTG TCACTCAATC 8640

AAATTGAAAA AGTAGTGGAT GATGGAATTT ACGTCAAGGT CAAGTTTGGT ATTCGTCAGG 8700

AGGGACTTAT CTTTGTGCCG AACATGCAGC TTGATATGGA AGAGGATAAG GTGAAGGTTT 8760

TCATCAGGGA AACCAGCTCC TACTATGTCT ACCACAAAGA CGCTGCCGAG AAAAATTGTT 8820

ATATGAAAGG TCGAACCTTA ATTAGACAGT TCAGCTATGA AAATCAAACC ATTCCATTAC 8880

GCAGAAAAGC GACAGTCGAT ATGATTAAAG AGAAGATTGC GGAAGTGGAT GCTTTGATTG 8940

AACTGGAAGT AGAAAATCAA TCTTATGTCA CGATTAAAGA TGAGTTAGTG CATGAACTAG 9000

CAGCGTCTGA ATTGAGAATC AATGAGTTGC AAGAACGAAT GTCAACCTTG AATCAAGTAG 9060

CAGAATATCT ACTGGCTTCA GTTGAAAGTA AGCAAGAAAT GAAATTAAAT CTTTCAAAAC 9120

TGAATATAAC TGAGAATATC AGTGCTAATA TTGTTGAGAA AAAATTGAAG AGCCTGGGGA 9180

520
ATCAACTGGA ATTGGAAAGG GGCAGGTATG AAAAGATGGT AGT 9223

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6827 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

TCTGCTGGCT ACCATCATCT GACTTGGGCA AGACCAAAGT CTTAGTTACA ACTGTATTCT 60

TCTCAGCATT TTCAATAACT GGCAATGCCG ACTGAAGCGT ATCTTTTTCT GTTTTTGTAG 120

CTGGTCCAGT TTCTTTTTTC TGTCCGCAAC CAACCAGGAC AAAAAGGAAA GCTAGACTAA 180

CAAGAACTAT TTTTTTCATT TCTTTCTTCT TTCTTTTTGA AATTAAAATA GAATAAGACT 240

GGGAAGTGCT CCCAGCCTTG ATGTTTATAG AGCTGCACGC AAACGTGCTT CTGCATTTTC 300

TACATTACGG ACAGAGCGTG GTAGGAAGGC ACGAATATCG TCTTCCTTGT AGCCAACTTG 360

CAGGCGTTTT TCATCTACAA GGATTGGGCT CTTTAAAATT CTCGGTGTTT CCATAATCAG 420

ATTGAGAACT TCATTGACAC TCAAATCTTC AATATCCACT CCAAGGGCTT TGGCATAGCG 480

ATTTTTAGAC GAAACGATGC TGGCTATTCC GTTATCTGTT TTGGTTAGAA TATCCAGTAA 540

TTCTTCTCTC GTAATTCCTT CTTTACCAAG GTTTTGTTCT TTATAACTTA ACTGGTGGGC 600

ATTGAGCCAG GTTTTTGCTT TTTTACAGCT AGTACAACTT GAGACTGTAT AAATTTTAAT 660

CATGTACCTA CCCCTTTCGC TACATGTTAC TATCAGTTTA GTCTATTATA CCATAAAAAA 720

CATCCGACTT GCGACCTATT TTTAATTTTT TTTGACTTTT TTCGTCATTT TCGTACTTTT 780

TTCTTGACAA ACAACTAAAT GACTATCAAC TCTTTTGGAG CTAGGGTCAA TAATTCACAA 840

CCTGTCTCTG TAATCAGGAT ATCATCCTCG ATACGAACGC CATATTTGCC TTCGATATAG 900

ATACCTGGTT CATCGGTCAA GGCCATACCT GTCTTAATAG TTTCTGTAGA AGTCTGACTA 960

AAGTAGGGTT CCTCATGGAT ATCCAGACCA ATACCGTGGC CAATGCCGTG AGTAAAGTAG 1020

TCACCATAAC CTGCCTCAAT GATAATATCA CGAGGGATTT TGTCAAAGTC ACGGAAACCT 1080

AAGCCTGCCT TAGCTTGGTC AATCAAGGCT TGGTTAGCTT TTAGAACCGT ATTGTAAATC 1140

TCTGCCTGCT CATCGCTAAC ATGCCCTAGA TAGATAGTCC GGGTCATATC ACTGACATAG 1200

TGGTCATAGA GACAGCCGAA GTCCATGGTG ATGGCTTCTC CCAACTCCAC TGGTTTGTGC 1260

ATTGGATGGG CATGGGGTTT AGAAGAATTG ATACCGCTAG CTAGGATCGT ATCAAAAGAT 1320

AAGCCAGATG CTCCCAACTC ACGCATGCGG AAATCAAGGA AGTTGGCAAT CTCAATTTCA 1380

GTTTTTCCTG GTTTGATAAA GTCAAGCGCA TCGCGGAAAG CTTGGTCTGA GATAGAACAA 1440

GCCTTGCGAA TCGCTGCAAT CTCTGCCTCA TCCTTAATCA TACGAAGACC TTCCACAAAC 1500

TGAGTTTGTG GAAGCAAGTT CAAACCTGCA AAAGCTGCCT GCATACGGTG GTAATAAGAC 1560

ACTGAAATCT CATCTTCAAA ACCGATACGA GTCAAGCCCA TGTCCTTAAC AATTCCTGCA 1620

ATGACAGCCA ATTCATCACG ATCAGCCACA ATCTCAAAAC CACTGGTTTC TTGCTTAGCT 1680

GCGATGATAT AGCGAGAGTC TGTCACTAAG ACCTGACGGT CACGACTGAT AAAGACTGTT 1740

CCGTTTGAGC CCCAAAAACC AGTCAAATAA TAGACGTTTT TAAGATTGTT GATGATGATA 1800

CCATCTAGTT CTTTTTCTTG CATTTTAGCT AGAAATGCTT GTACGCGTTT ATTCATGATG 1860

TAACTTTCCT TTCAAATAGT GTCCTGTATA GCTGGCTTCG TTGGCAGCTA CTTCTTCTGG 1920

AGTTCCTGTT ACGATGATGG TTCCACCACC GACACCGCCC TCAGGTCCCA AGTCAATGAT 1980

ATGGTCTGCC GTCTTGATAA CATCCAGATT GTGCTCGATG ACGAGGACTG TATTGCCATC 2040

GTCTACAAAG CGAGCTAAAA CCTTGAGCAG GCGAGCAATG TCCTCTGTAT GAAGCCCTGT 2100

CGTCGGCTCA TCCAGAATGT AGAAAGATTT TCCTGTCGAT CGTTTGTGGA GTTCGCTAGC 2160

TAACTTCATA CGTTGGGCTT CTCCCCCAGA AAGGGTGGTA GCTGGCTGTC CCAAGGTCAC 2220

ATAGCCTAGC CCTACATCCT TGATGGTCTG GAGTTTGCGT TGAATTTTCG GAATGTGTTG 2280

GAAAAATTCT ACCGCATCGT TGACCGTCAT ATCCAAGACC TGCGAAATAT TCTTTTCCTT 2340

GTAGTGAACT TCTAGGGTTT CACTGTTATA GCGGGTTCCG TGGCAAACTT CACAAGCCAC 2400

ATAAACATCT GGCAAGAAGT GCATCTCAAT CTTGATAATC CCGTCACCTG AGCAAGCTTC 2460

ACAGCGACCT CCCTTGACGT TGAAACTGAA GCGCCCCTTC TTGTAGCCTC GAATCTTGGC 2520

TTCATTTGTC TGAGCAAAAA GGTCACGTAT ATCGTCAAAA ACTCCTGTAT AGGTAGCTGG 2580

GTTAGACCTC GGCGTCCGTC CGATAGGGCT CTGGTCAATA TCAATCAAAC GGTCGACATG 2640

CTCAATCCCT GTAATAGTCT TAAACTTACC AGGTTTGTCT GAATTACGGT TGAGCTTCTG 2700

GGCAATGGCT TTTTTGAGAA TGCTGTTGAT TAGAGTCGAT TTCCCTGAAC CCGACACACC 2760

TGTCACTGCG ATAAATTTTC CTAGTGGAAA GCGAGCCGTG ACATTTTGCA AGTTGTTCTC 2820

ACGCGCTCCT ATCACTTCAA TAAAACGACC ATTTCCGACA CGGCGCTCTT CTGGTACTGG 2880

GATGACACGT TTGCCTGACA AGTACTGACC TGTGATAGAC TTGCTGTTGC GAGCCACTTG 2940

CTTAGGTGTA CCTGCTGCAA CAATCTCACC ACCAAAAACA CCGGCACCAG GACCAACGTC 3000

AATCAGATAA TCAGCCTCAC GCATGGTATC TTCGTCGTGT TCCACCACGA TAAGAGTATT 3060

GCCCAAGTCA CGCATCTTTT TCAGACTGGC AATCAGGCGA TCATTGTCCC TCTGGTGAAG 3120

ACCGATTGAC GGCTCGTCTA GGATATAGAG GACACCTGAT AGGTTGGAAC CAATCTGGGT 3180

TGCCAAACGA ATGCGCTGAC TTTCCCCACC TGAAAGGGTT CCTGCTGAAC GTGACAGGGT 3240

TAGATAGTTA AGACCCACAT TATTAAGGAA GGTCAAACGA TCCTTGATTT CCTTGAGAAT 3300

GGGACGAGCA ATGATGGCTT CATTTTCAGA CAAAGTTAAC TGGCTCACCA AGTCCAAGTG 3360

GTCAGCGATA GACAGGTCTG AGATTTCTCC AATATGTGGC CCTTGCTGGC CGCCCACACG 3420

GACAGACAAG GCCTGGTCAT TGAGACGATA GCCTTGACAG GTTCCGCAGG TCAGCTCATT 3480

CATGTAGAGA CGCATCTGAG TGCGAGTGTA ATCGCTATTG GTTTCATGGT AACGACGTTT 3540

GATATTATTG ATAACTCCCT CAAACGGAAT GTCGATATCG CGCACGCCAC CAAATTCATT 3600

CTCATAGTGG AAATGGAATT CCTTACCATC TGACCCATAG AGAATCAAGT TCTTATCTTC 3660

TTCTGACAGG TCCTCAAAAG GCTTATCCAT AGCCACTCCA AAGACTTTCA TGGCCTGCTC 3720

TAACATGTTT GGATAGTAGT TGGATGAGAT AGGATTCCAA GGTGCTAGCG CTCCCTCACG 3780

TAAGGTTTTG CTAGCATCTG GCACTACCAA ATCAGTATCC ACCTCCAGCT TGATGCCCAA 3840

GCCGTCACAC TCACTACAAG AGCCAAAAGG AGCATTGAAA GAAAAGAGAC GAGGCTCTAA 3900

CTCTGGGACA GTAAAACCAC AAACTGGACA GGCATAATGC TCAGAGAACA ACAACTCCGA 3960

GTCGTCCATG GTGTCGATAA TGACATAACC TTCTGCAATA CGAAGGGCAG CCTCAATGGA 4020

ATCAAAGAGA CGACTACGAA TGCCCTCCTT GATAACAATA CGGTCAACCA CGACATCGAT 4080

ATTGTGTTGC TTGCTCTTAG ACAACTCTGG CACTTCGGTC ACATCATAGA CTTCCCCATC 4140

CACACGGACA CGAACATACC CGTCTTTCTG AACCTTCTCG ATAACACTCT TATGTTGGCC 4200

TTTTTTCTTG CGGATGACAG GAGCCAAGAT CTGCAAGCGC TGGCGTTCAG GTAACTCCAA 4260

AACCTTATCA ACGATTTGCT CCACAGAAGA AGCATTGATA GCTCCATGTC CGTTGATACA 4320

GTAAGGCGTC CCCACACGTG CGTAGAGGAG ACGCAGATAG TCATTGATTT CAGTCGTCGT 4380

TCCCACCGTC GAGCGAGGAT TTTTACTAGT CGTTTTCTGG TCGATGGAAA TAGCTGGGCT 4440

GAGACCATCA ATGGCATCTA CATCTGGTTT TTCCATATTT CCCAAGAACT GACGAGCGTA 4500

GGCGGACAAA CTCTCTACAT AGCGACGTTG TCCCTCCGCA TAGAGAGTAT CAAAAGCCAG 4560

ACTGGACTTC CCTGAACCTG ACAAGCCAGT CACGACAACC AACTTGTCTC GCGGAATCTC 4620

CACATCAATA TTTTTTAAAT TATGGGCACG CGCCCCATGA ATGACAATTT TATCTTGCAT 4680

CTTTGTTCTT TCTAGTCCAT TATTGCTTAC CATTATACCA AAAAAAGTGA GATTCTATTA 4740

CCCAAAAGGC CGATTTTGTA GTATAATAGT ACAGTGTGAA AAAATCTGAA AAATGAGAAA 4800

GGATAAGGGA TATGAAACAA GTTTTTCTCT CTACAACAAC TGAATTTAAA GAGATCGATA 4860

CGCTTGAACC GGGTACTTGG ATCAATCTCG TCAATCCGAC TCAAAATGAA TCACTCGAAA 4920

TCGCCAACAC CTTCGATATT GATATTGCTG ACCTTCGAGC ACCGCTCGAT GCGGAAGAAA 4980

TGTCTCGTAT TACCATTGAA GACGAGTATA CCCTGATTAT CGTAGACGTG CCGGTCACGG 5040

AGGAAAGAAA TAACCGCACC TACTACGTAA CCATCCCGCT TGGTATTATC ATCACTGAGG 5100

AAACCATTAT CACTACGTGT TTGGAACCAC TACCTGTCCT TGATGTCTTT ATCAACCGTC 5160

GATTGCGTAA TTTCTATACC TTCATGCGTT CACGTTTTAT CTTTCAAATT CTTTATCGCA 5220

ATGCAGAGCT TTACCTAACA GCCCTTCGTT CAATCGACCG CAAGAGTGAA CAAATCGAAA 5280

GTCAACTGCA TCAATCAACT CGTAATGAAG AATTGATTGA GCTCATGGAA TTGGAAAAAA 5340

CTATCGTCTA TTTCAAGGCC TCCCTCAAAA CAAATGAGCG CGTGATTAAG AAATTGACCA 5400

GTTCAACCAG CAATATCAAG AAATACCTTG AGGACGAAGA CCTGCTTGAA GACACCCTGA 5460

TTGAAACCCA ACAGGCCATC GAGATGGCAG ATATTTATGG AAACGTCTTG CATTCTATGA 5520

CAGAGACCTT TGCCTCTATC ATTTCTAACA ACCAGAACAA CATCATGAAA ACCTTGGCCC 5580

TTGTGACCAT CGTCATGTCC ATCCCAACCA TGGTCTTTTC TGCCTACGGG ATGAACTTTA 5640

AGGATAATGA AATCCCCCTA AACGGAGAGC CAAATGCCTT CTGGTTAATC GTCTTTATCG 5700

CCTTTGCTAT GAGTGTCTCG CTCACTCTCT ATCTCATCCA TAAAAAATGG TTCTAAGAGG 5760

AGTTCCTATG TCTCAAATTG ATCTACAAAA ATtAACTAAG AAAAACCAAG AGTTTGTCCA 5820

CATTGCTACC CAACAATTCA TCAAAGATGG GAAAACAGAC GCTGAAATCC AGACTATTTT 5880

TGAGGAAGTC ATTCCCCAAA TCCTTGAGGA GCAATCTAAA GGTACAACTG CCCGTTCCCT 5940

ATACGGCGCA CCAACTCATT GGGCTCATAG CTTCACTGTC AAAGAGCAGT ACGAAAAAGA 6000

GCATCCAAAA GAAAATGATG ACCCAAAACT GATGATTATG GACTCAGCTC TTTTCATCAC 6060

TAGCCTCTTT GCCCTTGTCA GCGCCCTCAC AACCTTCTTT GCGGCAGACC AAGCTTTCGG 6120

CTATGGATTG ATTACTCTTC TATTAGTTGG ACTGGTTGGT GGATTTGCCT TCTACTTGAT 6180

GTACTACTTT GTTTACCAAT ACTATGGACC AGATATGGAT CGCAGTCAAC GTCCACCTTT 6240

CTGGAAATCT GTACTAGTTA TCCTAGCTTC TATGTTCCTT TGGTTGCTTG TCTTCTTTGC 6300

AACAAGCTTC CTACCAGCTA GCCTTAACCC AGTACTGGAT CCATTGCCAC TAGCTATTAT 6360

TGGAGCAGCC CTCCTAGCCC TTCGCTTCTA TCTCAAGAAA CGCTTGAATA TCCGTAGTGC 6420

AAGTGCAGGA CCAACACGCT ATCAAGAATA AGAAAACGAT AAAAGCAACT GCAGGTGCGG 6480

TTGCTTTTTC ACTTACTTTT TTGAGTTATA TTCAATGAAA ATCAAAGAGC AAACTAGGAA 6540

GCTAGCTGCA GGTTGCTCAA AGCACAGCTT TGAGGTTGCA GATAAAACTG ACGTGGTTTG 6600

AAGAGATTTT CGAAGAGTAT TAAAAGTATT CTTCTGAAAT CCCACATAGC TTTCTCTTAT 6660

ATTTTGTGAT AAAATAGGCT CAATCTATTT CTAGGAGGAT GAGATATGGT TTCTACTATT 6720

GGTATTGTTA GTTTATCTAG TGGCATTATC GGAGAGGATT TTGTCAAACA CGAAGTGGAC 6780

TTGGGTATCC AACGTCTCAA GGATCTGGGA CTCAATCCCA TCTTTTT 6827

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11864 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

CTGGCTAGTT GCATAGAGCA AAGTTGCTTC TTCATCAACA AAACCGTTCA TTTCAAAATA 60

GGAAAGCAGC TCATCAGGAC TCTCCAAACG AATCCCTTTG TAATCCAGCT CAACTGCCAC 120

CTCTTTCAAG GCTGCAAGAA GAAGTGTTCC CAGGCCCTGT CTCTGATGGT CAAACTCGAT 180

GACTAAAGAA TGTACTTTTA GACATTGCGG ATTGTCTGAC TGGGGACTTG ATAAAATATA 240

GCCTAAAAGT TGATTTTCAT CCCTAGCTAG AAGAAAGGTA TCCGCACACT TACGGATACT 300

TTCTTCTAAA ATATGGGAAA GTTGCTGCTT TTCAGCTGGA AAAGACGAGG TCTGAAGTGC 360

CCCTATCTCA GGCAAATCAG ACTTGCTTGC CTGAATGATC TTAATTGGAA TTTCCATGGG 420

AACATCCTAT TGAACATTGC TTGTCAAGTT AGACAAGAGA CGCTCAAATG AGTATTCATA 480

GGTTTGGATG TCTCCTGCTC CCATAAAGAC GTAAACAGCA TTGTCATGGT CTAGGAGTGG 540

AGAAACATTT TCAACAGTAA TCACTTGGTG TTTTTTGTTG ATTTTGTTGG CTAGGTCTTC 600

TACCTTAACG TCACCATGAT CTACTTCACG AGCCGAGCCA TAAATTTGCG CTAGATAAAC 660

AGCATCTGCT TGGTTTAAAG CATCGGCAAA GTCGTCCAAC AAGGCAATGG TTCTTGTAAA 720

GGTATGCGGT TGAAAGACTG CTACAATTTC CTTGCTTGGG TATTTCTGAC GAGCCGCATC 780

CAAGGTCGCA ATAATTTCTG TTGGATGGTG GGCAAAGTCA TCGATAATCA CTGTATCATT 840

GACAATTTTC TCAGTGAAAC GACGTTTAAC ACCGGCAAAT GTTTTCAAGT GCTCACGCAC 900

CAAGTTCAAA TCAAATCCTG CTGTGTAAAG AAGACCAATA ACGGCTGTCG CATTCATGAT 960

ATTGTGACGA CCAAAGGTTG GAATGTGGAA TTGCCCCAAG TTTTGTCCAC GGAAATGAAC 1020

GGTGAAGGTT GAACCAGTTA TTGAACGAAG AAGATCACTA GCTACAAAGT CATTGCCTTC 1080

AGCTTCAAAA CCATAATAAT AAATTGGTGC ATCAGACGTA ATCTTACGCA ATTCAGCATC 1140

TTCACCATAG ACAAAAAGAC CCTTGGTGAT TTGTTTGGCA TAGTCGTTAA AGGCATTAAA 1200

AACATCCTCG AGACTTGTGA AATAATCTGG ATGGTCAAAG TCAATGTTGG TGATAATAGA 1260

GTATTCTGGG TGGTAAGGCA TGAAGTGACG CTCATATTCG TCAGATTCAA AGACAAAATA 1320

TTTGGCATTG GCCGAACCAC GACCTGTCCC ATCTCCAATC AAGAAGCTGG TATCTGTAAT 1380

GTGAGACAAG ACATGAGACA ACATACCTGT CGTTGAAGTT TTTCCATGTG CTCCTGCTAC 1440

TCCCATGCTA ACAAAGTCAC GCATAAAGCT ACCTAGAAAC TCATGGTAAC GTTTGTAGCT 1500

GATACCATTT TGGTCCGCAT AGGCAATTTC GACGTTGTTA TCTGGACGAA AGGCATTTCC 1560

AGCGATAATT TCCATATCAC CGTCTAGATT TTTTTCATCA AAAGGAAGAA TGGTAATTCC 1620

TGCCTGCTCA AGACCGCGTT GGGTAAAGTA GTACTTTTCA ACATCTGATC CCTGAACCTT 1680

GTGCCCCATC TGGTGCAACA TCAAGGCCAA GGCACTCATC CCTGATCCCT TAATTCCGAT 1740

AAAATGATAT GTCTTTGACA TGTTTTCTCC CCTATTCTGT CATTCTGGTC AGATTCAACT 1800

CTTGGGCAAC CCGACGTTCT TGTTCTGTTT GTTTACTTTT TTTATTGTAG ATTTGGCTCT 1860

TCTTTAGAAA ATCATAATTG TTTTTCTTTG GAGCAGGTGC TGACACTTCT TCATTCTTGG 1920

TAGGGATAGA ATGAACTTCT TCCGCCAAGA TATAATGAGA CTGGGTCAAT TTTTGGCTAT 1980

ATTTGACAAA TTCACCAGGA TTTTCCTTTT GGAAAGGAGC TGTCGGTTGA TTGCCCTGTC 2040

TAACTAGACT GGGCTGAGAA TGACGTCTCG CAAGGCTGAA ATCCTGAGTT AGGTAGTTAG 2100

CAGAGCGTTT CTTTTTCAAG TCCGCACGCG CTTCTTCACG CGCCACCTCC GCATAGCTCT 2160

TTCCTTCTTT TTTAACCCCT AAAGGAGCCT TTTTAGGTTT TTCGACTTGC TTTTCAATCG 2220

GTTTTACTGG TTTTTCTTCA GCAATAGGAG CCCATTCTAA ATAATTTTTA TCTCGATACT 2280

CACCCTTGAT ATTACTGATC AGATCAGACT CATCATAGAG ATTCATGACT GGCATTTCAG 2340

TCAACATGAC CTCGTCATCT GACACCAATG GAAATCGTTC TTGTTTCATT TTCTATTTCC 2400

TTTCAACACT TCATTATAGC GTATTGTCTT GATTTTTCAA GTGCTGGCTT CAGAAATTCC 2460

CAAAATTTCT CTAATTTCTG CTAGGGTCAG ACTACCACGT GACTCTGTGC CGTCCAATAC 2520

TTGTGACACC AGATGTTTCT TTTGTTCTTG GAGTTCCTGA ATTTTTTCTT CAATGGTTCC 2580

CTTGGTCACC AAGCGATAGA CCTCAACCGT TTCTTCCTGA CCCATCCGAT GGGCACGGCC 2640

AATGGCTTGC GCTTCCACCG CAGGATTCCA CCAAAGGTCA ACCAAGATCA CTGTATCTGC 2700

ACCTGTCAGG TTCAGACCGA CCCCACCAGC CTTGAGGGAA ATCAGAAAGG CATCTCTTTC 2760

TCCTTGGTTA AAGGCCTTGG TCATGTCTTG TCTTTCCTTG GCTGGGGTTG AACCCGTAAT 2820

TTTAAAGGAA GTCAGGCCCA AGTCTGGCAG TTCTTGTTCA ATTTTTTCCA ACATTCCCTT 2880

GAACTGAGAG AAAATCAAGA CACGGTGTCC GCCGTCTGCC ACCTGTACCA GTAGGTCTCG 2940

GAGACTATCT AGTTTGCCGC TGGCTCCCTG ATAATCTTCC ATAAACAGGG CAGGAGTGTC 3000

ACATATTTGA CGCAAGCGCA TCAAACCAGA TAAAATTTCC ACACGACTTC GCTGAAATTC 3060

CTGTTCTGAC ACTTGAGCCA GATGGTCTCG CATCTGTTGT AACTGGGCAA GGTAAATAGC 3120

CTTTTGCTGG TCTTCCAGTT CATTTTTATA AACCACCTCA ATCAAGTCTG GCAATTCAGT 3180

CAGAACTTCT TCTTTCTTGC GTCGCATCAC GAAAGGCTTG ATAAACTGAG CCACTCGCTC 3240

TGCTGGCAAT TTCATAAATT CTTTCTTGCT TGGCAAAAGT CCAGGCATGA CGATTTGGAA 3300

AATAGACCAC AACTCACCCA GATGGTTTTC AATCGGAGTT CCTGACAAGG CAAAGACCGA 3360

CGGCACCACA AATTGTCTCA AGGTCTGGGC AATCTTGGTC TGGGCATTTT TCATGACCTG 3420

AGCCTCATCT AAGAAAAGGA AGTCAAAGGC CATCCCTTGA TAAAACTCAC TGTCCTGACG 3480

GAAGGTGGCA TAGCTAGTCA CATAGATTTG ATGGCTCTCG GCAAGAATCT CCTCACGACT 3540

TGCTTTCAAA CCATGAACAA CAGTCACATC CAACTGTGGA GCAAATTTCT GAAACTCATC 3600

TGCCCAGTTG TAAATCAAAC CCGACGGAGC GAGAATCAAA ACCCGACTTT CTTTTGTCAC 3660

TTGACTAGTC AAAAAAGCAA TGGTCTGAAG GGTTTTCCCA AGTCCCATAT CATCAGCCAA 3720

AATCCCACCA AAACCATAAT GATGGAGCAT CTGCAACCAG CCAATTCCCT TTTCCTGATA 3780

ATCTCGCAAG TCAGCCTTGA CCTGAGTTGC TTGCAAAGGA AAGTCCTCTG GATGCGTCAA 3840

ATCCTGGGCC AGATTCTGGA ATTCTTGTGA AAAAGAAACA CGGTCTCGCC CTTCAAAGAG 3900

ATGAGCTAAA CTGTAGGCCA AGGATTTCCG AGCCTGCAAG GTCCCATCTT TTAATTCAAA 3960

TTGCCCCAGT TCCTGTAGAT TTTGGCGAAT TTTCTTGGTT TCTTCATCGA AAAAGTAAAC 4020

TTGATTAGAC GAATCAATAT AAAAATCCTG ATTGGCAACC AAGGCCTGCA TGGCTTGGTC 4080

GATTTCCTCC TGGACAATAT TTTGAAAATC AAACTGGATT TCCAAGAGAC CTCCCTTGGA 4140

GGCAATCTGC ACCTGAGGAC TCGCTAGGCT ATAAAGCTCT TCTAGTTTAT CTGATAGGTC 4200

AACATGCCCG AGTTTTTCAA AGACTGGAAT GATATCATGA AAAAAATGAT AGACAGACTC 4260

CGCTTTTAAG GCCTGACGCC AAGATTGAAA ATCGGCCTCA AAGCCCGCAG CCAAACAGAC 4320

TTGGAAAATT CTTTCTTCTA AGTCTGCGTC ACTTGAAAAG GGTAATTCTT CTAGCTCTTG 4380

TCGGCTAGAT ACCTGTCTAT TTCCATAATC AAACTGAATT TCTAAACGAA TCCGATTATC 4440

TTCTTCCCTG TCAAAGTAAA AAGAGGGCGC AAAAGTTTTG ATTTGTAGAC GTTCTGGAGC 4500

TGAAACGGTG CCCATCTGGA TAAAAAGAGT CAGACAGGAG GCCAATTTGT CTCGATCACT 4560

GCTATCAAAT TGCAGGTATT TCTTTCCTTG TTGACCCACA GGTAACGCTT TAATTTCCTT 4620

GAGAAGACGC ATCTGCTGGT CTGTTAAAAA ATAAACCTGA CCTTTATGGA AAAGTACTGC 4680

TCCCTGATAA AAGACATTGA CCCTAGGACT CTCACTGATT TCCATTTCAA AATAATCCGA 4740

GTATTCTGTT ACTGTAAAGG CAAATAGATT GGCATCAGCA TGCATATCCT GAAAAAGCAG 4800

GGTTTGGTAG CTATCCACTT GATGGTCAAA TTGAAAATGG GGCAAGGCCA TCAGTAAATT 4860

CACACCCTGC TCAAAAAAGG TCAGAGGGAA AAAGAGGTGC CGACCTTGGT TTTGGAAAAA 4920

GAGGTCTGGA ACCAGCCCTT CCTCCGTTAG TCCGTGCAAG AAAGTCAAAA GTTCTTGGCT 4980

GGCATCATCA AAGGCTTCCC AAGAAAGAGA CTCCTCATAA ATCTTGCCAA TCATATACGA 5040

CTTTCTCTGC TCGACAATCC TTAAAAAAAG TGGAATATCC CGAATGACAT AGTATTTTTG 5100

GCTATTGATT TGGCCGATTC TCAGAGTCCA CAAGATATGA TTGGTTCCTG CTTCCACCTG 5160

ACCCACAGCT GATAACTCAT AGGCGCATTC TGATTTTGGA GATAAAATTC GATCCAAAAA 5220

CTTGCCACCC AAGGTCACCT TGGTTTCAAC AGCCTCTTTT TCTTCATGAC CTTCTTCCAG 5280

ACTCCACAAG ATTTCCTGAC CACGCTCATC ATTTTTCAGA AAATGCTCTA GCGCTGCCAA 5340

ATGCACACAG TAGCCCCTCT TTTGAAAAAA ATCACAGGCA CAAAAAACCA AATCATCCTC 5400

TAAACTATAG CGCAGTTCTT CTTCTGCAAC GCGAGCGTAG AGCCGATTGT TCTTTTCCTT 5460

GATGATATCA ACCTTACCAG TTTCATAAAG GGCAACACCT TCGATACGAA TTTTCCCCGG 5520

AATCAATTTA GCCATATTTT CACCTTTACC TTATCTTTTT ATTATACCAT ATTTTCGCCT 5580

ATGAAAATAG CCTTCTAGGA AGACTTTTCT CCTAGAAGGC TGGATTTTTA ACGTTTGGCA 5640

AAAGTAGCCA CAATCCGCTG ACAGACTTCT TGCAACAGAG ATTTGGGCAT AGCTATATTG 5700

ATGCGGGCAT GGAGACTTCC TTCCTCTCCA AAATCCAAAC CACGGTTGAG GATAACCTTG 5760

GCTTCATTTC TCAACAACTC TTGCAATGTT TCATCAGTCA GGTCATAAGC TGAAAAGTCA 5820

AGCCAAATCA AGTAGGTACC TTGCGGTTTC ATGACCTTGA TTTTAGTCTC TTTTCCAAAT 5880

AGATCCATCA CATAATTGAT GTGGTCTTCA AAGACTTGCT TGAGTTCCTC TAGCCAATCT 5940

TTACCGTATC GATAGGCAGC TTCTGTCGCC AAATAACCCA AGCCTGAAAT TTCATGCTGA 6000

TTATTGGCCA ACAGGCGTTT CTGGAAAGCC AGTCTCAACT TAGGATTTTC AATGACTGCA 6060

TAGGAATTTT TTGTTCCAGC AATATTAAAT GTTTTAGTGG CACTGCTCAA GACGATAGCA 6120

AAATTTTTGA AGGCAGGATT GATGGTATTG AAAGACTGGT GTTTGTGACC AAAGAGGGTC 6180

AAATCTTGGT GAATCTCATC CGAAACTAAC AAAACACCGT GTTTTTGGCA GAGTTGGCCA 6240

ATCTTCTCCA ACACTTCTTT TTCCCAAACA CGTCCACCAG GATTGTGAGG GTTGCAAAGA 6300

ACATAGAGTT TAACCTCCTC TTCCACCAAA TCCTTTTCAA GTTGGTCAAA GTCAATCTCA 6360

AACAGACTAT CCTTTTCCAC TAAGGAATTA GTAATCAATC TACGATTATT CAACTTGACA 6420

CTGCGAGCAA AGGGTGGGTA GACAGGCGTG TTAATTAAAA CCGCCTCGCC TTCTTTTGTA 6480

AAGGTTTGAA TAGCTGTTGA GATGGCTGGT ACCACACCCT CGATAAAGAC AAGAGCCTCT 6540

TTGTCAAAGT TGTAACCGTA TTGTGTAGCT TCCCACTTTT GAACTTCCTT AATTAAGTCT 6600

TCACTGGCAT AGGTATAACC ATAAACCAGT TGGTCTGCGT AAGTTTGCAC GGCTTGGCGG 6660

ATTTCAGGCA AGACCACAAA GTCCATATCC GCTATCCAAG CTGGTAGAAC TTCACTATCC 6720

GTTTCTGTTT CTTTCCATTT ATAGGTATGG TGCCCTAAAC GGTTGGGCAG GCTTGTAAAA 6780

TCATATTTTC CCATCTTTGT CTTATCCTTC TATGGCTTGG CGCAAATCTG CAATCAAATC 6840

TCTAGCATCC TCAATCCCAA TAGACAAACG CAAGAGGTCA TCTGTCAAAC CATAAGAATG 6900

GCGTACCTCT GCTGGAATAT CAGCATGAGT TTGAGTCGTT GGATAAGTAA TAAGACTTTC 6960

CACTCCACCC AAACTTTCCG CAAAAGAGAA GACCTTGAGA CTGTTCAAAA TATGAGGAAT 7020

GCGTGTTTCA TCGGCTACTT TAAAGGAAAT CATGCCTCCA CGACCAGTGT AGAGAACTTC 7080

CTTAACTGCT GGAGAATCCT TCAAAAAGGC AACCACTTCT TGGGCGTTAG CTGTTGAGCG 7140

CTCCATACGA AGAGACAAGG TCTTGAGACC ACGAAGCAAC TGGTAGCTGT CAAATGGAGA 7200

CAAGACTGCC CCTGTTGTAT TAAGATTGTA AAAAAGCTTC TCGTATAGTT CTAAACTATT 7260

GGTCACAACC ACTCCAGCCA AGACATCATT GTGGCCTGCT AGATACTTGG TTGCTGAATG 7320

GAGAACGATA TCTGCTCCAT CTTCAATCGG ACGTTGGTAG ATAGGGCTAT AGAAGGTATT 7380

GTCCACCACC ACTTTGGCAC CCTTAGCATG AGCCAATTTT GCTAGTTTTT CGATATCAAA 7440

TTCCAACATC AAGGGATTGG TTGGGGTTTC GATATAGAGA ACATCCACAT CCTTTTCTAA 7500

CTCGGCAATC AACTCTTCTT CTGTATTGGC ATAGGTAAAA TGGAAATGAC CTTCCTGCTC 7560

CACTTGGTTA AACCAGCGAA AAGAACCACC GTAAAGATCA CGCACTGCCA AGACCTTACT 7620

TCCTACTGGA AAGACGCTAA AGGCCAGTAC AATAGCTGAC ATCCCTGAGC TAGTCGCTAG 7680

GGCATAGTCT GCTGACTCAA TAGCCGCCAA GACTTCCTCA GCCTTACTAC GAGTTGGATT 7740

TTTAGTGCGC GTATAGTCAA ACCCAGTAGA TCGACCAAAC TCTGGATGCT GATAGGTCGT 7800

TGAAAAATGA AGTGGTGTCA CCAAAGCACC TGTTGCCTCA TCAGACTTGA TCCCTGCTTG 7860

TGCTAAAATT GTGTTAATGT GTAATTCCTT GCTCATACAA TTCCTCCAAA TCTATAGTAA 7920

CTATTGTACC ACTTATTTTG TATCCTTCGT TTTCTTGTTT TCAAGAGCTA GTTATAGTTT 7980

CAAACTATAT AAAAAGGGAG TTTTTCCTGC TCCCTTTAAT AGACTATAAA ATGGTGAATC 8040

TCAAAAGACA CCTTCACTCT ATCATTTGCT CCTGCACAAA ACGAGCATAA CGCTCATGAT 8100

TTTCCAGTAG TTCCTTATGA GTTCCTGAGC CAGTGATTTT CCCCTCCTCT AAGAAGAAAA 8160

TACAATCCAC ATCTTTTACC GTTGACAAAC GATGCGCTAT AATCACAACC GTCTTCTCCT 8220

TTAGTACAGA ATAGAGGCTA CTGATAATCG CATACTCAGA ATCCGCATCA AGATTAGCAG 8280

TGGCTTCATC AAATATAAGA ATTTCAGCAT CTTTTAAGTA GGCTCTAGCT ATTTGAAGTC 8340

TTTCGTTCGC CCCCCTGACA AGAGTCGTCC GCGTTCACCA ACTTCAGTAT CTAGTCCCTC 8400

TTTCATGGAG CGAATCTCAT CACCTAGTGA TACTAAGTCT AGCACTTTCA TCAATTCATC 8460

ATCAGTTACT AAGCGATTCA AACCGAGACA AAGATTGTCA CGAATACTGC CAGATAAGAC 8520

TGCATTATTT TGTGAAACCC AAGCGATTTT ACTTCTCCAT TCTTTTAAGT TAAAATCATA 8580

TATACTTGAT TGCTCCATTA GAATATCTCC TGAAAGCGGT TTATAAAACC GCTCTAACAA 8640

ACGCACAATC GTTGATTTTC CTGATCCAGA TGGTCCAACA AAAGCAATTT TTTGCCCCTT 8700

GAAAATTGAA CAAGTAATAT CCTTTAAGAC AGGTCGATTT TCATCATAAC CAAAATAGAC 8760

ATGGTTAAAA TTCAACCCTC GTCCTGATAC CGATTTTCCT CCCTCAAATT TTTCTTTAGG 8820

AACTGCAAGC AAGTTCTCCA GTGCAACTGA AGATCCCTTG CTCCTAGAAT AAACAGTTAC 8880

AAAATTAGCT ATATTACTAA TAGGATTAAG TAATTGAAAG AGGTAAATCA AAAACGAAAC 8940

CAAGGTTCCC ACAGATATAT ATCCTGCGCT GACCCGATAA CCCCCATAGG TTAGCATCAC 9000

AGCTATAGTC GCAAAGATAA ATAAGAGAGC AAACGGGGTC TCAAAAGAAG TAACCCTATC 9060

TGATTTCAGT GAATTGTTTT GTACCCTTTC AATACAATTA TCCAAAAACAT CCTGTACACT 9120

TTTCTCTGCT TGGTTAGTCT TAATTAATTC ATGTTCTTGA ATCTTTTCAG TCAATTGCCC 9180

TGTTAAATTT CCTCCTGTAA ACGACGACTA TACTTTTCAC TGATATTGGA AAGGGGCAAG 9240

ATAATAAACA TCATACAAGG AAGAGTGATG AATAAAAGTA GAGAAAGATT CCAATCAAGA 9300

CTAAATAAGA CTACAATGGA ACCAAGTACC ATAACTAAAC TCAGAATAAT ATTTGGGAAA 9360

GTCGTAATTA AAAACTCACG AATGACACTC GTGTCATTGA CAATGGCAGA AGTCAACTCC 9420

CCACTTTGGC TCTTATCAAA GAAGGATTTC TCTACATAAA TCAACCCCTC TATCACTTTT 9480

TTCCTGATTT TTGCTATCTT TTTTTCACCC GATTGACTAA ACAGATAGTA ACCAATAGAA 9540

GAAAACAAGG CTTGACCAAT AAAAATCAAA AACGATTGAA ATACTTTGGA GCCTATATTT 9600

TCAATAGAAC TCCCATCTAT TAAATCCTTT AAGATAAGGG GAAGCAACAA AGCAAGTAGA 9660

CTAGACAGAA CAAGTAAGAA ACTCCCCATA ATCACCTTAG TATCTACTCT TAATAATTTT 9720

AATTTCATAA ATACTCCTTA TAATATTTCA ACGGATAAAG TCGGGAATAA CTCAATTTGA 9780

GGATAAAATC TAATAAATCT TCCTATAACA AAACGCATAA CATCTAGGAT TTTATATACC 9840

TGATATTATG CGTTTTTAAG CACAAAGACT TCTTACACAA ACTTATCTAC AATTAGATTT 9900

TATTTGACAT GTTTTGCCAA TTCTTCTTGG GCTTTTTTAT TGGATTCTTC TTTTTCTTTC 9960

AACCATTTTT CTCTGGCTTT TGCATATTCG TCTGTTGTGA CAATCTTATC TTGTACTTTG 10020

AGGTATTTAT ATGATTCAAC CCCTTTTGTA CCGGTTAAAC CATAGGCAGC AGCAAATGGT 10080

ACGGTTCTTC TCAATGATGG TGTTCCCCCA CGCGAAACAC TTGGAAGAAC TAAAGAACTA 10140

TCAATCAACC AAGCTTGAAT ATCAGCATAT TTCTCATAAC GTTTGGCCGG ATCTTGCTCT 10200

TTATTAGCTT CTTCCAACAT TTGAGTATAG ACATCCAGTC CAACTGCCTT AGCCTTGTCA 10260

TTGGCCTCAC CAGGCTCTAG TCCAAGATTT TGCAGAAATC CTCCACTATT AGTATTAAAA 10320

ATATCGAGAT AGGTTGACGG GTCTTGATAA TCAGGTCCCC AACCGCCATG ATATAAATCA 10380

TAATCTTTCT GAGCAGCTGT TTGAGCAAAG TAGCCTGAAC TGTCAAACTC ATCTGATGTT 10440

AATTGCTGAA TGTCAATCAC TACATTATCA GAACCTAAAA CAGATTCAAT TGATTGTTTG 10500

ATAGAACTAA CTCCTTGTAT GCCTACTTTA TCTGTTACTT CCACAGTCTT ATCCAAGTGG 10560

ATTGGGAATT GAACACCCTT TGCTTCGAGT TCTTTCTTAG CTTCCGCAAA CTTAGCCTTG 10620

GCTTTCTCAG GATTGTAGTA AGGGTCTTGA CCATCCGCAA AGTTGATACC TTGCCATTCC 10680

TTACCATAGT TGACCATCTT AGAGGCTACA ACTTCACCAA AGTCTTTTCC CTTGATACTG 10740

ACAAAGTTTG GAGGAACCAC TAGGTTACGC AAAATCTTTG TTGCACCTTC TTTCCCTTCA 10800

GACTGAGCCC CATAAGATGT TCTGTCAAAA GCAAAATTGA TAGCCTGACG GAAGTTTTTA 10860

TTGAGAACTG CTTCCTGAGT CGATTTCTTT TCAATGTCAC TTGTTTTAGA AGTATAATTG 10920

TAAGACTTCC TATCTAGGTT AAAATTAAAG AAATATGAAG TTGAATTTTG CATACTATAG 10980

ATGATATTGT TTTTGTATTT TTCTTTAATC CCTTCATAGC TGGAGCTGTT AGGAAAAAGA 11040

CGAGCCGTAG TATAAGCACC AGCTGTAAAA TTACGTTCCA GTGATTCTTG GTCGCTACCA 11100

TCATAGTAGG TCAATTTCAC ATCGTCTACA AAGACATTCT TAGCATCCCA GTAATTAGGG 11160

TTTTTCTTAT ATTCAATAGC AGATTTTGAG ACAAGTGCTT TCATCAAGAA AGGTCCATTG 11220

TACAAAATAC TAGATGGATC CGCCTTCCCA AAATCATCCC CTTTTGATTT CAGGAAATCT 11280

GCATTAACAG GAAAAAGTAT CGTTGCAAGT GTTTTTGAAT TCCAGTAAAG TTCTGGTTTA 11340

ACCAAAGTAT ATTGAACCGT TTGGTCATCA AGTGCCTTGA CACCGACAGT TGAAAAGTCG 11400

CTTGTTTTAC CAGTGATATA GTCATCCAAA CCAGCAACAG AGTCCTGCAC TAGATACAAG 11460

GCTTCTGATT TTTTATCAGC TGCATATTGC AAACCTGTCA CAAAATCCTG GGCAGTTACA 11520

GGCGCATATT CTTCTCCCTC AGAAGTAAAC CACTTGGCAT CCTTACGAAG TTTGTAGGTA 11580

TAGGTCAAAC CGTCCTGAGA AACAGTCCAA TCCTCTGCTA ATGATGGAAT AATATTCCCA 11640

TATTGGTCAT TTTCTAATAA CCCGTCTACC AAATTTGCAA CAATATCGGA TGTTGCTGCG 11700

CGGTTTTCTG CTAGATAGTT CAAGCTAGAT GGATCACTTG AATAAACATA GTTGTAGGTT 11760

TTTGACGCCG TGCTAGAATT TCCACACGCG CTCAATAAAA CTCCTGTACC CAGGACAAGA 11820

CCTGCCAAGG TTAGATATTT GCTCTTAGAC TTTTTCATTT CCGG 11864

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2412 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

TAACTGCACT AAACATAATA TAAGGAGAGA AAATGTCTGC AATAGAACGT ATTACAAAAG 60

CTGCTCACTT AATTGATATG AACGATATTA TCCGTGAAGG GAATCCTACT CTACGCGCGA 120

TTGCTGAGGA AGTCACTTTC CCCCTATCTG ACCAGGAAAT CATCCTAGGC GAAAAGATGA 180

TGCAATTCCT TAAACATTCC CAAGATCCTG TCATGGCTGA AAAAATGGGA CTCCGCGGTG 240

GTGTTGGACT GGCTGCTCCC CAGTTAGATA TCTCAAAACG CATTATCGCT GTTTTGGTAC 300

CTAATATTGT TGAAGAAGGC GAAACTCCAC AGGAAGCCTA CGATTTGGAA GCCATTATGT 360

ACAATCCAAA AATCGTCTCT CACTCTGTTC AAGATGCTGC TCTTGGCGAA GGAGAAGGTT 420

GCCTGTCTGT TGACCGTAAC GTGCCTGGCT ATGTTGTTCG CCATGCCCGC GTTACTGTTG 480

ACTACTTTGA CAAAGATGGA GAAAAACACC GTATCAAACT CAAAGGCTAC AACTCCATTG 540

TTGTTCAGCA TGAAATTGAC CACATTAACG GTATCATGTT TTACGATCGC ATCAATGAAA 600

AAGACCCATT TGCAGTTAAA GATGGTTTAC TGATTCTTGA ATAAAGAAAA TCCCGTTGCA 660

AGACGGGGTT TTGTGTTATA ATAGAGGCAT GAAAACAAAT GATATTGTCT ATGGTGTCCA 720

CGCCGTTACC GAAGCCCTCC TTGCAAATAC AGGAAACAAA CTCTACCTCC AAGAAGATCT 780

CCGAGGTAAG AATGTTGAGA AAGTCAAGGA ACTAGCTACA GAAAAGAAGG TGTCCATTTC 840

TTGGACATCA AAAAAATCTC TCTCTGAGAT TACTGAAGGT GCTGTTCATC AAGGTTTTGT 900

TCTACGAGTG TCTGAATTTG CCTATAGCGA GCTAGATTAC ATCCTTGCAA AAACACGCCA 960

AGAAGAAAAT CCACTTCTAT TGATTCTAGA TGGTCTAACC GATCCCCATA ATCTGGGTTC 1020

TATCTTGCGA ACAGCCGATG CGACCAATGT TTCAGGTGTC ATCATTCCCA AGCACCGTAC 1080

TGTCGGAGTA ACTCCTGTCG TTGCCAAAAC AGCCACAGGT GCTATTGAAC ACGTtCCAAT 1140

TGCCCGAGTG ACCAACCTCA GTCAAACCTT AGGATAAACT TAAGGATGAA GGTTTCTGGA 1200

CCTTTGGAAC GGATATGAAC GGTACTCCTT GCCACAAGTG GAATACAAAA GGGAAAATCG 1260

CCCTCATCAT TGGAAATGAA GGAAAAGGTA TCTCTAGCAA CATCAAAAAA CAGGTCGATG 1320

AAATGATTAC CATTCCGATG AATGGACATG TTCAAAGCCT TAATGCCAGT GTTGCTGCGG 1380

CCATTCTCAT GTACGAAGTT TTCCGAAATA GACTATAAAA AAGTTTCCAG TCATCTGATT 1440

GGAAACTTTT TTATGATTAA CTATGTTCTG TAATGAATTT ATAGGCTTCT TGACCAGCGA 1500

TAGCTCCATC TCCAACCGCT GTTGTTACTT GGCGAAGGTC TTTCAAGCGA ACATCTCCAA 1560

CTGCAAAGAT ACCGTCGACT GCAGTTTTCA TGTGGTTATC TGTCACAATC CATCCTGCCT 1620

GATCTTGGAT ATTCAATTCT TTAACAAAAT CGCTAAGAGG GTCCAAACCA ACATAGATAA 1680

AGACACCACC GAAGGCTTGT TCTGTCACTT GACCTGTTTT CACATTTTCA AATACGACTG 1740

ATTCTACTCG GTTTTCACCC TTGATTTCCC TTACTACAGA ATCCCAGATA AAGCTGATTT 1800

TTTCATTCGC AAAGGCGCGA TCTTGTAAAA CCTTTTGGGC ACGAAGTTGG TCACGACGGT 1860

GAACAATGGT AACAGTCTTA GCAAAACGAG TCAAGAAGAG GGCTTCTTCA ACAGCTGAAT 1920

CTCCACCACC AACTACCAAT AAATCTTGGT CACGGAAGAA AGCACCATCA CACACAGCAC 1980

AGTAAGAAAC ACCACGACTG TTCAGTTCTT CTTCTCCAGG CACTCCCAAA GGACGGTGTT 2040

TAGAACCAGT TGCTACGATA ACTGTACGTG TTTCATATGT TTGGTCATCA GTCATCACTT 2100

TCTTAAAATC ACCATGGCTT CGACATTTTC AACATAACCA TAAATGTGCT CAACACCAAG 2160

ATTTTCAAGT GGTTCAAACA TCTTTTCAGC CAATTCAGGT CCACTAATAT TAGCGTATCC 2220

TGGGTAATTT TCGATATCAG ATGTATTATT CATCTGACCA CCTGGCAGAC CACCTTCAAT 2280

CAAAGCTACT TTTAGATTGC TTCGAGCAGC ATACAAGGCC GCAGTCATCC cTGCAGGTCC 2340

AGCACCGATA ATAATAGTAT CGTACATATA GATTCCTTCT TTCTTGGTGT AACTATCTTT 2400

ATTCTAACTC TG 2412

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7760 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

CCGATTTGGT GGAATTTTTG TCTCATCATT TAGAAGGTGT TGCAAGAGCA GAGTTTACCT 60

TGGTGCTTCA TACCAAATTG GGAGAAGCCT CTGTTTTGGC AAATATTGTA GATGTAAACA 120

AGGATGAATG GATTTTAGGA ACAGTTGCTG GTGCCAATAC CTTATTGGTT ATTTGTCGAG 180

ATCAGCACGT TGCCAAACTC ATGGAAGATC GTTTGCTAGA TTTGATGAAA GATAAGTAAG 240

GTCTTGGGAG TTGCTCTCAA GACTTATTTT TGAAAAGGAG AGACAGAAAA TGGCGATAGA 300

AAAGTTATCA CCCGGCATGC AACAGTATGT GGATATTAAA AAGCAATATC CAGATGCTTT 360

TTTGCTCTTT CGGATGGGTG ATTTTTATGA ATTATTTTAT GAGGATGCGG TCAATGCTGC 420

GCAGATTCTG GAAATTTCCT TAACGAGTCG CAACAAGAAT GCCGACAATC CGATCCCTAT 480

GGCGGGTGTT CCCTATCATT CTGCCCAACA GTATATCGAT GTCTTGATTG AGCAGGGTTA 540

TAAGGTGGCT ATCGCAGAGC AGATGGAAGA TCCTAAACAA GCAGTTGGGG TTGTTAAACG 600

AGAGGTTGTT CAGGTCATTA CGCCAGGGAC AGTGGTCGAT AGCAGTAAGC CGGACAGTCA 660

GAATAATTTT TTGGTTTCCA TAGACCGCGA AGGCAATCAA TTTGGCCTAG CTTATATGGA 720

TTTGGTGACG GGTGACTTTT ATGTGACAGG TCTTTTGGAT TTCACGCTGG TTTGTGGGGA 780

AATCCGTAAC CTCAAGGCTC GAGAAGTGGT GTTGGGTTAT GACTTGTCTG AGGAAGAAGA 840

ACAAATCCTC AGCCGCCAGA TGAATCTGGT ACTCTCTTAT GAAAAAGAAA GCTTTGAAGA 900

CCTTCATTTA TTGGATTTGC GATTGGCAAC GGTGGAGCAA ACGGCATCTA GTAAGCTGCT 960

CCAGTATGTT CATCGGACTC AGATGAGGGA ATTGAACCAC CTCAAACCTG TTATCCGCTA 1020

CGAAATTAAG GATTTCTTGC AGATGGATTA TGCGACCAAG GCTAGTCTGG ATTTGGTTGA 1080

GAATGCTCGC TCAGGTAAGA AACAAGGCAG TCTTTTCTGG CTTTTGGATG AAACCAAAAC 1140

GGCTATGGGG ATGCGTCTCT TGCGTTCTTG GATTCATCGC CCCTTGATTG ATAAGGAACG 1200

AATCGTCCAA CGTCAAGAAG TAGTGCAGGT CTTTCTCGAC CATTTCTTTG AGCGTAGTGA 1260

CTTGACAGAC AGTCTCAAGG GTGTTTATGA CATTGAGCGC TTGGCTAGTC GTGTTTCTTT 1320

TGGCAAAACC AATCCAAAGG ATCTCTTGCA GTTGGCGACT ACCTTGTCTA GTGTGCCACG 1380

GATTCGTGCG ATTTTAGAAG GGATGGAGCA ACCTACTCTA GCCTATCTCA TCGCACAACT 1440

GGATGCAATC CCTGAGTTGG AGAGTTTGAT TAGCGCAGCG ATTGCTCCTG AAGCTCCTCA 1500

TGTGATTACA GATGGGGGAA TTATCCGGAC TGGATTTGAT GAGACTTTAG ACAAGTATCG 1560

TTGCGTTCTC AGAGAAGGGA CTAGCTGGAT TGCTGAGATT GAGGCTAAGG AGCGAGAAAA 1620

CTCTGGTATC AGCACGCTCA AGATTGACTA CAATAAAAAG GATGGCTACT ATTTTCATGT 1680

GACCAATTCG CAACTAGGAA ATGTGCCAGC TCACTTTTTC CGCAAGGCGA CGCTGAAAAA 1740

CTCAGAACGC TTTGGAACCG AAGAATTAGC CCGTATCGAG GGAGATATGC TTGAGGCGCG 1800

TGAGAAGTCA GCCAACCTCG AATACGAAAT ATTTATGCGC ATTCGTGAAG AGGTCGGCAA 1860

GTACATCCAG CGTTTACAAG CTCTAGCCCA AGGAATTGCG ACGGTTGATG TCTTACAGAG 1920

TCTGGCGGTT GTGGCTGAAA CCCAGCATTT GATTCGACCT GAGTTTGGTG ACGATTCACA 1980

AATTGATATC CGGAAAGGGC GCCATGCTGT CGTTGAAAAG GTTATGGGGG CTCAGACCTA 2040

TATTCCAAAT ACGATTCAGA TGGCAGAAGA TACCAGTATT CAACTGGTTA CAGGGCCAAA 2100

CATGAGTGGG AAGTCTACCT ATATGCGTCA GTTAGCCATG ACGGCGGTTA TGGCCCAGCT 2160

GGGTTCCTAT GTTCCTGCTG AAAGCGCCCA TTTACCGATT TTTGATGCGA TTTTTACCCG 2220

TATCGGAGCA GCAGATGACT TGGTTTCGGG TCAGTCAACC TTTATGGTGG AGATGATGGA 2280

GGCCAATAAT GCCATTTCGC ATGCGACCAA GAACTCTCTC ATTCTCTTTG ATGAATTGGG 2340

ACGTGGAACT GCAACTTATG ACGGGATGGC TCTTGCTCAG TCCATCATCG AATATATCCA 2400

TGAGCACATC GGAGCTAAGA CCCTCTTTGC GACCCACTAC CATGAGTTGA CTAGTCTGGA 2460

GTCTAGTTTA CAACACTTGG TCAATGTCCA CGTGGCAACT TTGGAGCAGG ATGGGCAGGT 2520

CACCTTCCTT CACAAGATTG AACCGGGACC AGCTGATAAA TCtACGGTAT CCATGTTGCC 2580

AAGATTGCTG GCTTGCCAGC AGACCTTTTA GCAAGGGCGG ATAAGATTTT GACTCAGCTA 2640

GAGAATCAAG GAACAGAGAG TCCTCCTCCC ATGAGACAAA CTAGTGCTGT CACTGAACAG 2700

ATTTCACTCT TTGATAGGGC AGAAGAGCAT CCTATCCTAG CAGAATTAGC TAAACTGGAT 2760

GTGTATAATA TGACACCTAT GCAGGTTATG AATGTCTTAG TAGAGTTAAA ACAGAAACTA 2820

TAAAACCAAG ACTCACTAGT TAATCTAGCT GTATCAAGGA GACTTCTTTG ACAATTCTCC 2880

ACTTTTTTGC TAGAATAACA TCACACAAAC AGAATGAAAA GGAGCTGACG CATTGTCGCT 2940

CCCTTTTGTC TATTTTTTAA GGAGAAAGTA TGCTGATTCA GAAAATAAAA ACCTACAAGT 3000

GGCAGGCCCT GGCTTCGCTC CTGATGACAG GCTTGATGGT TGCTAGTTCA CTTCTGCAAC 3060

CGCGTTATCT GCAGGAAGTC TTAGGCGCCC TCCTTACTGG GAAATATGAA GCTATTTATA 3120

GTATCGGGGC TTGGTTGATT GGTGTGGCCG TAGTCGGTCT AGTTGCTGGT GGACTCAATG 3180

TTGTCCTCGC AGCCTATATT GCCCAAGGAG TTTCATCCGA CCTTCGGGAG GATGCCTTCC 3240

GTAAAATTCA AACCTTTTCT TATGCTGATA TTGAACAATT TAATGCGGGA AATCTAGTCG 3300

TTCGAATGAC AAATGATATC AACCAGATTC AGAACGTTGT CATGATGACC TTCCAAATTC 3360

TTTTCAGACT TCCCCTCTTG TTCATCGGTT CGTTTATCCT AGCGGTTCAA ACCTTACCTT 3420

CTCTGTGGTG GGTGATTGTT CTCATGGTAG TCTTGATTTT TGGTTTGACT GCTGTCATGA 3480

TGGGAATGAT GGGGCCTCGT TTTGCCAAGT TTCAAACCCT TCTTGAGCGC ATCAATGCCA 3540

TTGCCAAGGA AAATTTACGT GGCGTTCGTG TGGTCAAGTC CTTTGTCCAA GAAAAAGAGC 3600

AATTTGCTAA GTTTACAGAG GTCTCAGACG AGCTTCTTGG TCAAAACCTT TACATTGGTT 3660

ATGCCTTTTC AGTAGTGGAA CCCTTTATGA TGTTGGTTGG TTACGGGGCG GTCTTCCTCT 3720

CTATTTGGCT GGTCGCGGGA ATGGTTCAGT CGGATCCGTC TGTTGTTGGT TCCATCGCTT 3780

CTTTTGTTAA TTACCTAAGC CAGATTATCT TTACCATTGT TATGGTTGGA TTTTTGGGAA 3840

ATTCTGTCAG CCGTGCCATG ATTTCCATGC GTCGTATTCG AGAAATTCTT GACGCAGAGC 3900

CAGCTATGAC CTTCAAGGAT ATCCCAGATG AAGAGTTGGT TGGAAGTCTT AGCTTTGAAA 3960

ATGTGACCTT TACCTATCCA ATGGACAAGG AACCGATGCT GAAAGATGTG AGCTTTACTA 4020

TTGAACCTGG TCAAATGGTT GGTGTAGTTG GAGCGACTGG TGCAGGAAAG TCAACCTTGG 4080

CTCAATTGAT TCCACGTCTC TTTGATCCAC AGGACGGGGC CATTAAAATC GGTGGCAAGG 4140

ATATTCGAGA AGTGAGTGAA GGAACCCTGC GTAAAACAGT TTCCATCGTT CTCCAACGTG 4200

CCATTCTTTT TAGTGGAACG ATTGCAGATA ACTTGAGACA GGGGAAGGGG AATGCTACTC 4260

TATTTGAAAT GGAGCGCGCA GCCAATATTG CCCAGGCTAG TGAATTCATT CATCGTATGG 4320

AGAAAACCTT TGAAAGTCCA GTTGAAGAAC GGGGAACCAA TTTCTCTGGT GGACAAAAAC 4380

AAAGGATGTC GATTGCGCGT GGGATTGTCA GCAATCCACG TATTCTGATT TTTGATGATT 4440

CGACCTCAGC CTTGGATGCC AAATCAGAGC GCTTGGTGCA AGAAGCTTTG AATAAGGACT 4500

TGAAGGGGAC GACAACCATT ATTATTGCTC AAAAAATTAG CTCGGTTGTC CATGCAGACA 4560

AGATCTTGGT TCTAAATCAA GGACGATTGA TTGGTCAAGG TACGCATGCA GACTTGGTTG 4620

CCAACAATGC CGTTTACCGT GAAATCTATG AAACACAGAA ATGAAAGACA AACTATAAGA 4680

AAAGTCAATA GTTTTATCTA AACTATTTCT TATTTCAATT TGATGATTTG GCGATGATTT 4740

TAGAGCACGG CAAAAAGCCC TTGAAAAAGT CCATTTTTTC AAAGGTAATC CTGTGTTAAT 4800

TTCAGAAATT ACATCACTTT TTGTTCGTCA AATGGCAGCT CTTTTTTTAG GATATAAAAC 4860

AGGGTTCGGA TAAGTTTTTT TGCAAGGTGG ATGATGGCTA CATTGTAATG TTTTCCTTGT 4920

TCTAATTTAG TCTTAAGATA GGCCTTAAAA GCAGGCGAAA AGCGAGGGCA TGCTTTGGCA 4980

GCTTGTATGA GTACCTACCG CAGATGAGGG GAACTCCGTT TGACCATTCT TCCTGCTAAA 5040

TCAATCTGAT CTGACTGATA AATAGAAGAA TCCAGTCCAG CGAAAGCTTG TAATTGAGCA 5100

GGATTATCAA AGGCATGAAT ATTTCGAATC TCAGCTAAAA TGACCGCCCC TAAACGATCC 5160

CCAATCCCAG TAACCGTCGT GATGACCGAG TTGAACTCAG CCATCAAGTC ATTGACACAT 5220

GTTTCCGCCT TGTCAATGAG CCTCTTGTAA TGTTTGATGT TTTCATTACA CGAGATAAAA 5280

CGTCTATGCG TTATCAAACT CATTACCAAT TAAAACAAAA AGCTGTGGTT AGATCCTTTC 5340

GGAAATTGTC AAGCGATTGG AGGAAATGAA CTAATCCACA GCGGCTTATT CCAAGTATAC 5400

CACTTGGGCT TTGGCAGTAG CTAACTGCGC TAAATATAAT ATAAGGAGGA GTAAAATGAA 5460

GACAGTTCAA TTTTTTTGGC ATTATTTTAA GGTCTACAAG TTCTCATTTG TAGTTGTCAT 5520

CCTGATGATT GTTCTGGCGA CTTTTGCCCA AGCCCTCTTT CCAGTCTTTT CTGGACAAGC 5580

GGTGACGCAG CTAGCCAATT TAGTTCAAGC TTATCAAAAT GGCAATCCAG AACTTGTATG 5640

GCAAAGCCTA TCAGGAATCA TGGTCAATCT TGGCCTGCTG GTTTTGGTTC TATTTATCTC 5700

TAGTGTAATA TACATGTGTC TCATGACGCG CGTGATTGCA GAATCGACCA ACGAGATGCG 5760

CAAAGGCCTC TTTGGTAAGC TTGCTCAGTT GACGGTTTCT TTCTTTGACC GTCGACAAGA 5820

TGGCGATATC CTGTCTCATT TTACCAGTGA TTTGGATAAT ATCCTCCAAG CCTTTAACGA 5880

AAGCTTGATT CAGGTCATGA GCAATATTGT TTTATACATT GGTCTGATTC TTGTCATGTT 5940

TTCGAGAAAT GTGACGCTGG CTCTCATCAC CATTGCCAGC ACCCCATTGG CTTTCCTTAT 6000

GCTGATTTTC ATCGTGAAAA TGGCACGCAA ATACACCAAC CTCCAGCAGA AAGAGGTAGG 6060

GAAGCTCAAC GCCTATATGG ATGAGAGCAT CTCAGGCCAA AAAGCCGTGA TTGTGCAAGG 6120

AATTCAAGAG GATATGATGG CAGGATTTCT TGAACAAAAT GAGCGCGTGC GCAAGGCAAC 6180

CTTTAAAGGA AGAATGTTCT CAGGAATTCT TTTCCCTGTC ATGAATGGGA TGAGCCTGAT 6240

TAATACAGCC ATCGTCATCT TTGCTGGTTC GGCTGTACTT TTGAATGATA AGTCTATTGA 6300

AACAAGTACA GCCCTAGGTT TGATTGTTAT GTTTGCACAA TTTTCACAGC AGTACTACCA 6360

GCCTATTATC CAAGTTGCAG CGAGTTGGGG AAGCCTTCAG TTGGCCTTTA CTGGAGCTGA 6420

ACGAATTCAG GAAATGTTTG ATGCAGAGGA GGAAATCCGA CCTGAAAAGG CTCCAACCTT 6480

CACTAAGTTG CAAGAAAGTG TTGAAATCAG TCATATCGTT TTTTCATACT TGCCTGATAA 6540

ACCTATTTTG AAAGATGTCA GCATTTCTGC CCCTAAAGGC CAGATGACAG CAGTTGTTGG 6600

GCCGACAGGT TCAGGAAAAA CGACTATTAT GAACCTCATC AATCGCTTTT ATGATGTTGA 6660

TGCTGGTGGT ATTTATTTTG ATGGTAAAGA CATTCGTGGC TATGACTTAG ATAGTCTTAG 6720

AAGCAAGGTG GGAATTGTAT TGCAAGATTC GGTCTTGTTT AGCGGAACGA TTAGAGACAA 6780

TATCCGATTT GGTGTGCCAG ATGCTAGTCA GGAAATGGTT GAGGTAGCAG CAAAAGCAAC 6840

CCACATTCAC GACTATATCG AAAGTTTGCC TGATAAGTAC GATACTCTTA TTGATGATGA 6900

CCAGAGCATC TTTTCAACAG GGCAGAAGCA ATTGATTTCA ATCGCTCGAA CCCTGATGAC 6960

AGATCCAGAA GTTCTCATTC TCGATGAAGC AACTTCAAAC GTAGATACGG TGACAGAAAG 7020

CAAGATTCAG CATGCCATGG AGGTGGTTGT AGCAGGTAGA ACTAGTTTCG TCATTGCCCA 7080

CCGCTTGAAA ACCATTCTCA ATGCAGATCA GATTATTGTC CTTAAAGATG GAGAAGTCAT 7140

TGAACGTGGT AACCACCATG AACTTTTGAA GCTAGGTGGC TTTTATTCAG AACTCTATCA 7200

CAATCAATTT GTTTTCGAAT AAGAAAGAAG TTGTCCTATG TGGGCAGCTT TTTCTTGTCC 7260

ATAAAAAATG TTTATCACAG CCTTAAAAAA AACATATTAG ACGAAAGTCA TTTTGAGTGA 7320

TATGATAGGA CTATCGTTAG CATTCGAAAG GAGAGGCATC ATGGCTAGAA CGGTTGTAGG 7380

AGTTGCTGCA AATCTATGTC CCGTAGACGC AGAAGGCAAA ATCATTCATT CATCTGTATC 7440

TTGTAGATTC GCAGAGATCA TTCGTCAAGT CGGTGGTCTC CCTTTAGTCA TTCCTGTTGG 7500

TGATGAGTCA GTTGTACGTG ATTATGTGGA AATGATTGAC AAACTCATTT TGACAGGAGG 7560

CCAAAATGTT CATCCTCAGT TTTATGGAGA GAAAAAGACC GTCGAGAGCG ATGATTACAA 7620

TCTGGTCCGT GACGAATTTG AATTGGCACT CTTGAAGGAA GCGCTTCGTC AGAATAAACC 7680

AATTATGGCA ATCTGTCGCG GTGTCCAACT TGTCAATGTT GCCTTTGGTG GAACCCTCAA 7740

TCAAGAAATC GAAGGTCAGG 7760

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2723 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

GAGGTTTTAA TTCACTTACC TCTsCCGTAT CTTTATTTAA AATGAATTCT TTTACGGTTG 60

TATTTCTTGC AAAATCTTTT ACAACAATCT TAATGTTTAG TGTCTTGTCT ATTATTTGTT 120

TAATATCATT AAATGATGTA TATTCTTTTC CATTTATATA AATATGTTGT TCTTGAATCT 180

CACCATCGAA TCCATTATTT CTTTTATCAT TGATGTTAAA GACTACAGAT TTTCCATCAG 240

CATATTCGAT ACTAGTATTT CCCTTAGGAT CAATGTTTAC TTCGGGTTTA ACATTATCAT 300

ATAAAAACTG ATAGTGGACT CCAACTGCTT TAGCATTCAA ATCGCTATAG CCAGTTTGAA 360

GATAAACATT TCCATCCATA TCTGTTACCT TATCTGGAAA TCCGTTTGCT TTATAGTCTT 420

TCATTCCCCA GTCCATGATG TCACCGTCTT TAACATTCAG CTTAATATTA AAATCTCTAG 480

TGTTATCAAT GTGTAAATCT CCGTAGATTA AATAATTATC TACAACCGAT TCATTAACTC 540

TCAATTCCCA GTTAAAACCA CCCTTATCAG AAATCTTACC TCTTAAATAA AATTCTGGAT 600

TTCGTACATA AATTTTATTA GATTTAGATG GATTAAAGTA GTTCTTATCC ATTGAAAGGT 660

TTACTGGTTT GGTATCAATA AATAACATGG AGCCATCTTC TTTTATAGCT TCTACATTGA 720

ACTTATCCTC TCCAGTGTAT TCTTTATCAT CCTTACCAAA TAATACAAGT TTAGAAGAAT 780

CTGTCACAAG ATTTCCGTCT TTATCGATAG CTTCCCCTTT ATCGTTCATT TTAAATGTAA 840

ACACTTGATA CCTTATAATG TTAAAGCCGT CCAAAGCCGA CATTAATACA GATTGGGTAC 900

TTCTTCCATC TTCAACATTT CTACTATCAG CATAAATTGT TGTTTCTGAA AGGGCTCTTA 960

GATTAGGATT GGCCTTTTGT ATTTTTGCTA TATCTTCCTT GCTATAGACT CCATTTCCTT 1020

CTAACATATC CGTTTTTCCA GGATTATAGG TAGTCACTTT TAGTGCATAG CCTTTTCTTA 1080

GAATGATATT ATCCTTTAAC AGATATTGTT GTTTTTCTGA ATCAGAATAG ATTTTACCAG 1140

ATTCCATTTT AGTTAAATTG TCTGGTTTGT TTTTTGAAAG ATCTCCTTCC CCTAATTCTA 1200

TGACATTCCC ATAACTTGAT ACATAGGGAT ATTCTGATTT AGTTTCCTTA ATTTTTTCAG 1260

GCATTCTAAT TTTAATTTCA GCTTTTTTCT GATCATTATC TTTAACAAAT AATCTCATAT 1320

CTCCTGCAAA AGCTAATCCA TCCACAATAT CATTAATATT AGCGTATAGA TCAAATGTCA 1380

TCGTTTTTGA GTGGAAATCA TACTTGGTCG CTTTGATTTC TATAGATTTA TAGTTATTCC 1440

CATAATATAC CTTGGCATTT TTAGAAACAT TACTTATCTT TCCAAGAATT TCAAAGTGTC 1500

CATCTTTAGA CGGACTTAGA ACACCATAAA TTTTTGATTT GATTTCGTCA AGTTTCTCAG 1560

TTTCATATTC TAGATCAGTC CCATCATCGT AGGCTATTAT ATTTCCTTTA TCATCGTATT 1620

TATAATCGTA TTCCTCCATT CTCTTACCAG TTTCACTTGT AAAATCATCA ACTTCTCTAA 1680

ATTTCTTTTT AATGAGTTTC TTTAAGTCTT TATTTTCAAA GTCTCTAATT GTTGAAATAT 1740

TTCTATCAAT AGTAAAACTA GATTTTTCTT TAATAGACTC TTCATTTTCT TGATGATGAT 1800

GTTCTACCCC AGTTGTATCT TTTTTTAGAC TACCCTCTTT TCCATTTCCT AAATTTTTAA 1860

ATTTAGATTC TGCAATCTCG CCAAGCTTTT GATATTTAGA TGAATCTTGA TCAGGATCTA 1920

CTAGATAATA GGAAATCATC CCCTTTTCAT CAGCCTGATT AGCAAATTTA ATTCTATGAA 1980

TCTTTGTGAA ATTGCTAGAA CCATCTAATG CAATGACTTC AATGATTTTT CCCCTTAAAT 2040

CTCCCGCACC TTTAATTTCA TAAATGGTAT TTCCGTCTTT ATCAAGTTTT CTATTTCTTC 2100

CTTGACCCTC ACCTGCGTAA GTTACTTCAA GATTTTTTTC AACCTCTCCA TCTTCATTAA 2160

CAAGAGCGGC GCCAGCATAC CAAACTTCGT TCGCAATCTC GTCAAATTTT TCAGGATGTT 2220

CTTTTTGATC TCTCGCAAAT AGCGTTTCAT TCTTATACTG ATCTTTTACC TTATGATAAG 2280

TATCCTTTGT AATCAACTTA ATTTTTTCAG GATTTGAAAA ATCAACCGAA ACAATCTTAG 2340

GGGCGGTGTT ATCAATTTTT ACAGGAATAT AGGAAACCTG CCATGGGTAA TCTTTAGTTA 2400

ATCTATATTT AAATTTATAG AAATATTGAC CTTCCGCAAT CGGTTCAAAT TGACCTCTTA 2460

TCTTAGTAGC AGGATCTTGA TTATCCTTAC TTTCTGGTGC ATTTTCTTCT CTACCTCTAG 2520

GATTATAGAT GAGTCCATCC CACTTCAAGT CACCCCAAAC TTTTAGTTTA GATGATTTGA 2580

TTCCCTTTGC ATCATTGCTT TTAGAATTTA AAATTCCTCT AATAAAGTGT TCTCTCGAAA 2640

TGACTTTTAA GTCTCTTTGA TTTTCTCCCT CTTTATTTGT ATTTACTATT GAAATCAATC 2700

CTTCTTCTGC ACTTCTTAAT ACA 2723

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11831 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

AAAAAAGTGG GAATGACTCA AATCTTCACT GAAGCTGGCG AATTGATCCC TGTAACAGTT 60

ATTGAAGCAA CTCCAAACGT TGTTCTTCAA GTTAAAACTG TTGAAACAGA CGGATACAAC 120

GCTATCCAAG TTGGTTTCGA TGACAAACGC GAAGTATTGA GCAACAAACC TGCTAAAGGA 180

CATGTAGCGA AAGCTAACAC GGCTCCTAAG CGCTTCATTC GTGAATTCAA AAACGTTGAA 240

GGCTTGGAAG TTGGTGCTGA AATTACAGTT GAAACATTCG CAGCTGGAGA CGTTGTTGAC 300

GTAACGGGTA CTTCTAAAGG TAAAGGTTTC CAAGGTGTTA TCAAACGCCA CGGACAATCA 360

CGTGGACCAA TGGCTCACGG TTCTCGTTAC CACCGTCGTC CAGGTTCTAT GGGGCCTGTT 420

GCACCTAACC GCGTATTCAA AGGTAAAAAC CTTGCAGGAC GTATGGGTGG CGACCGCGTA 480

ACAATTCAAA ACCTTGAAGT TGTACAAGTT GTTCCAGAAA AGAACGTTAT CCTTATCAAA 540

GGTAACGTAC CAGGTGCTAA GAAATCTCTT ATCACTATCA AATCAGCAGT TAAAGCTGGT 600

AAATAATAAA GAAAGGGGAA ATCAGTCACA ATGGCAAACG TAACATTATT TGACCAAACT 660

GGTAAAGAAG CTGGCCAAGT TGTTCTTAGC GATGCAGTAT TTGGTATCGA ACCAAATGAA 720

TCAGTTGTGT TTGATGTAAT CATCAGCCAA CGCGCAAGCC TTCGTCAAGG AACACACGCT 780

GTTAAAAACC GCTCTGCAGT ATCAGGTGGT GGACGCAAAC CATGGCGTCA AAAAGGAACT 840

GGACGTGCTC GTCAAGGTTC TATCCGCTCA CCACAATGGC GTGGTGGTGG TGTTGTCTTC 900

GGACCAACTC CACGTTCATA CGGCTACAAA CTTCCACAAA AAGTTCGTCG CCTAGCTCTT 960

AAATCAGTTT ACTCTGAAAA AGTTGCTGAA AACAAATTCG TAGCTGTAGA CGCTCTTTCA 1020

TTTACAGCTC CAAAAACTGC TGAATTTGCA AAAGTTCTTG CAGCATTGAG CATCGATTCT 1080

AAAGTTCTTG TTATCCTTGA AGAAGGAAAT GAATTCGCAG CTCTTTCAGC TCGTAACCTT 1140

CCAAACGTGA AAGTTGCAAC TGCTACAACT GCAAGTGTTC TTGACATCGC AAATAGCGAC 1200

AAACTTCTTG TCACACAAGC AGCTATCTCT AAAATCGAGG AGGTTCTTGC ATAATGAATT 1260

TGTATGATGT TATCAAAAAA CCTGTCATCA CTGAAAGCTC AATGGCTCAA CTTGAAGCAG 1320

GAAAATATGT ATTTGAAGTT GACACTCGTG CACACAAACT TTTGATCAAG CAAGCTGTTG 1380

AAGCTGCTTT CGAAGGTGTT AAAGTTGCCA ATGTTAACAC AATCAACGTA AAACCAAAAG 1440

CTAAACGTGT TGGACGTTAC ACTGGTTTTA CTAACAAAAC TAAAAAAGCT ATCATCACAC 1500

TTACAGCTGA TTCTAAAGCA ATCGAGTTGT TTGCTGCTGA AGCTGAATAA TCTAAGGAGG 1560

AAATATCGTG GGAATTCGTG TTTATAAACC AACAACAAAC GGTCGCCGTA ATATGACTTC 1620

TTTGGATTTC GCTGAAATCA CAACAAGCAC TCCTGAAAAA TCATTGCTTG TTGCATTGAA 1680

GAGCAAGGCT GGTCGTAACA ACAACGGTCG TATCACAGTT CGTCACCAAG GTGGTGGACA 1740

CAAACGTTTC TACCGTTTGG TTGACTTCAA ACGTAATAAA GACAACGTTG AAGCAGTTGT 1800

TAAAACAATC GAGTACGATC CAAACCGTTC TGCAAACATC GCTCTTGTAC ACTACACTGA 1860

CGGTGTGAAA GCATACATCA TCGCTCCAAA AGGTCTTGAA GTAGGTCAAC GTATCGTTTC 1920

AGGTCCAGAA GCAGATATCA AAGTCGGAAA CGCTCTTCCA CTTGCTAACA TCCCAGTTGG 1980

TACTTTGATT CACAACATCG AGTTGAAACC AGGTCGTGGT GGTGAATTGG TACGTGCTGC 2040

TGGTGCATCT GCTCAAGTAT TGGGTTCTGA AGGTAAATAT GTTCTTGTTC GTCTTCAATC 2100

AGGTGAAGTT CGTATGATTC TTGGAACTTG CCGTGCTACA GTTGGTGTTG TCGGAAACGA 2160

ACAACATGGA CTTGTAAACC TTGGTAAAGC AGGACGTAGC CGTTGGAAAG GTATCCGCCC 2220

AACAGTTCGT GGTTCTGTAA TGAACCCTAA CGATCACCCA CACGGTGGTG GTGAAGGTAA 2280

AGCACCAGTT GGTCGTAAAG CACCATCTAC TCCATGGGGC AAACCTGCTC TTGGTCTTAA 2340

AACTCGTAAC AAGAAAGCGA AATCTGACAA ACTTATCGTT CGTCGTCGCA ACGAGAAATA 2400

ATATTAAACT AGTCGCTTAA GCAACTAGTA AATCCGCCAG CTCGGTAGCG CTCCATAGGA 2460

GTGCAAGCCG CTGTGGTACA ACATTTAAAG GAGAAAATAT AAAAATGGGA CGCAGTCTTA 2520

AAAAAGGACC TTTCGTCGAT GAGCATTTGA TGAAAAAAGT TGAAGCTCAA GCTAACGACG 2580

AAAAGAAAAA AGTTATTAAA ACTTGGTCAC GTCGTTCAAC GATCTTCCCA AGTTTCATTG 2640

GTTACACTAT TGCAGTTTAT GACGGACGTA AACACGTACC TGTTTACATC CAAGAAGACA 2700

TGGTAGGCCA CAAACTTGGT GAATTTGCAC CAACTCGTAC TTACAAAGGT CACGCTGCAG 2760

ACGACAAGAA AACACGTAGA AAATAAGGAG AACATAAATG GCAGAAATTA CTTCAGCTAA 2820

AGCAATGGCT CGTACAGTAC GTGTTTCACC TCGTAAATCA CGTCTTGTTC TTGATAACAT 2880

CCGTGGTAAA AGCGTAGCCG ATGCAATCGC AATCTTGACA TTCACTCCAA ACAAAGCTGC 2940

TGAAATCATC TTGAAAGTTT TGAACTCAGC TGTAGCTAAC GCTGAAAACA ACTTTGGTTT 3000

GGATAAAGCT AACTTGGTAG TATCTGAAGC ATTCGCAAAC GAAGGACCAA CTATGAAACG 3060

TTTCCGTCCA CGTGCGAAAG GTTCAGCTTC ACCAATCAAC AAACGTACAG CTCACATCAC 3120

TGTAGCTGTT GCAGAAAAAT AAGGAGGTAA AATCGTGGGT CAAAAAGTAC ATCCAATTGG 3180

TATGCGTGTC GGCATCATCC GTGATTGGGA TGCCAAATGG TATGCTGAAA AAGAATACGC 3240

GGATTACCTT CATGAAGATC TTGCAATCCG TAAATTCGTT CAAAAAGAAC TTGCTGACGC 3300

AGCAGTTTCA ACTATTGAAA TCGAACGCGC AGTAAACAAA GTTAACGTTT CACTTCACAC 3360

TGCTAAACCA GGTATGGTTA TCGGTAAAGG TGGTGCTAAC GTTGATGCaC TCCGTGCAAA 3420

ACTTAACAAA TTGACTGGAA AACAAGTACA CATCAACATC ATCGAAATCA AACAACCTGA 3480

TTTGGATGCT CACCTTGTAG GTGAAGGAAT TGCTCGTCAA TTGGAGCAAC GTGTTGCTTT 3540

CCGTCGTGCA CAAAAACAAG CAATCCAACG TGCAATGCGT GCTGGAGCTA AAGGAATCAA 3600

AACTCAAGTA TCAGGTCGTT TGAACGGTGC AGATATCGCC CGTGCTGAAG GATACTCTGA 3660

AGGAACTGTT CCGCTTCACA CACTTCGTGC AGATATCGAT TACGCTTGGG AAGAAGCAGA 3720

TACTACATAC GGTAAACTTG GTGTTAAAGT ATGGATCTAC CGTGGTGAAG TTCTTCCAGC 3780

TCGTAAAAAC ACTAAAGGAG GTAAATAACC AATGTTAGTA CCTAAACGTG TTAAACACCG 3840

TCGTGAGTTC CGTGGAAAAA TGCGCGGTGA AGCAAAAGGT GGAAAAGAAG TAGCATTCGG 3900

TGAATACGGT CTTCAAGCTA CAACTAGCCA CTGGATCACT AACCGCCAAA TCGAAGCTGC 3960

TCGTATCGCC ATGACTCGTT ACATGAAACG TGGTGGTAAA GTTTGGATTA AAATCTTCCC 4020

ACACAAATCA TACACTGCTA AAGCTATCGG TGTGCGTATG GGATCTGGTA AAGGGGCACC 4080

TGAAGGTTGG GTAGCACCAG TTAAACGTGG TAAAGTGATG TTCGAAATCG CTGGTGTATC 4140

TGAAGAGATT GCACGTGAAG CGCTTCGACT TGCTAGCCAC AAATTGCCAG TTAAATGTAA 4200

ATTCGTAAAA CGTGAAGCAG AATAAGGAGA AGGCATGAAA CTTAATGAAG TAAAAGAATT 4260

TGTTAAAGAA CTTCGTGGTC TTTCTCAAGA AGAACTCGCG AAGCGCGAAA ACGAATTGAA 4320

AAAAGAATTG TTTGAACTTC GTTTCCAAGC TGCTACTGGT CAATTGGAAC AAACAGCTCG 4380

CTTGAAAGAA GTTAAAAAAC AAATCGCTCG CATCAAAACA GTTCAATCTG AAGCGAAATA 4440

ATAGACTAGG GAAGGAGAAA TTTCAATGGA ACGCAATAAT CGTAAAGTTC TTGTTGGACG 4500

TGTTGTATCT GACAAAATGG ACAAGACAAT CACAGTTGTA GTTGAAACAA AACGTAACCA 4560

CCCAGTCTAT GGTAAACGTA TTAACTACTC TAAAAAATAC AAAGCTCATG ATGAAAACAA 4620

TGTTGCCAAA GAAGGCGATA TCGTACGTAT CATGGAAACT CGCCCGCTTT CAGCTACAAA 4680

ACGTTTCCGT CTTGTAGAAG TTGTTGAAGA AGCGGTCATC ATCTAATCAA ACCTGAAAGG 4740

AGAAAACTGA AATGATTCAA ACAGAAACTC GTTTGAAAGT CGCAGACAAC AGCGGTGCTC 4800

GCGAAATCTT GACTATCAAA GTTCTTGGTG GTTCAGGACG TAAATTTGCA AACATCGGTG 4860

ATGTTATCGT GGCATCTGTA AAACAAGCTA CTCCTGGTGG TGCGGTTAAA AAAGGTGACG 4920

TTGTTAAAGC AGTTATCGTT CGTACTAAAT CAGGTGCTCG TCGTGCTGAT GGTTCATACA 4980

TCAAATTTGA CGAAAACGCA GCAGTTATCA TCCGTGAAGA CAAAACTCCT CGCGGAACAC 5040

GTATCTTTGG CCCAGTTGCA CGTGAATTGC GTGAAGGTGG CTTCATGAAG ATCGTGTCAC 5100

TTGCTCCAGA AGTACTTTAA TTTTTAGGAA CAAACTAGTC CCCTAGCTTC AAGCTAGGGT 5160

GCCCTTATGG GCGTAAGAAA AATCAAGGAG AAACCTAATG TTTGTAAAAA AAGGCGACAA 5220

AGTTCGCGTA ATCGCTGGTA AAGATAAGGG AACAGAAGCT GTTGTCCTTA CTGCCCTTCC 5280

AAAAGTAAAC AAAGTTATCG TTGAAGGTGT TAACATTGTT AAGAAACACC AACGTCCAAC 5340

TAACGAGCTT CCTCAAGGTG GTATCATCGA GAAAGAAGCA GCTATCCACG TATCAAACGT 5400

TCAAGTTTTG GACAAAAATG GTGTAGCTGG TCGTGTTGGA TACAAATTTG TAGACGGTAA 5460

AAAAGTTCGC TACAACAAAA AATCAGGCGA AGTGCTTGAT TAATCACGAA GGAAAGGAGA 5520

AGTATAATGG CAAATCGTTT AAAAGAAAAA TATCTTAATG AAGTAGTTCC TGCTTTGACA 5580

GAACAATTCA ACTACTCATC AGTGATGGCT GTGCCTAAAG TAGATAAGAT TGTTTTGAAC 5640

ATGGGTGTTG GTGAAGCTGT ATCAAACGCT AAAAGCCTTG AAAAAGCTGC TGAAGAATTG 5700

GCACTTATCT CAGGTCAAAA ACCACTTATC ACTAAAGCTA AAAAATCAAT CGCCGGCTTC 5760

CGTCTTCGTG AAGGTGTTGC GATCGGTGCA AAAGTTACCC TTCGTGGTGA ACGTATGTAC 5820

GAATTCTTGG ATAAATTGGT ATCAGTTTCA CTTCCACGTG TACGTGACTT CCACGGTGTC 5880

CCAACAAAAT CATTTGATGG ACGCGGGAAC TACACACTTG GTGTGAAAGA ACAATTAATC 5940

TTCCCAGAAA TCAACTTCGA TGACGTTGAC AAAACTCGTG GTCTTGACAT CGTTATCGTA 6000

ACAACTGCTA ACACTGACGA AGAGTCACGT GCATTGCTTA CAGGCCTTGG AATGCCTTTT 6060

GCAAAATAAT ATAGGAGGTA AATCTAATGG CTAAAAAATC AATGGTACCT AGAGAGGCTA 6120

AACGCCAAAA AATTGTTGAC CGTTATGCTG AAAAACGTGC TGCATTAAAG GCGGCAGGGG 6180

ACTACGAAGG TTTATCTAAA TTACCTCGCA ACGCCTCACC GACTCGTTTA CATAATCGTT 6240

GTAGGGTTAC GGGGCGCCCA CATTCAGTTT ACCGCAAATT TGGTCTGAGT CGTATCGCTT 6300

TTCGCGAACT TGCGCATAAA GGTCAAATTC CTGGTGTAAC AAAAGCATCT TGGTAATTTA 6360

AGATATCAAG AGCGTCAAAA CTCCAAGTAA AAATAGGAAA CTTGACGAAG AAACTAAAGT 6420

TTCTAGGAAA GTTTATCTTT TTCACACAGA GTTTAGCCCG GGTTCAATTG GGCTTGCCAA 6480

TTTGAACACG AGCTACAGCT TTGGCAAAAA AGACCAATTT GCTTTGGAGC ATTGCTTCTG 6540

CATTAAATTG TCTATTTTTG CTCGTGCTGT TACGCTCTTT GTATCATGTA TTAACTAGCA 6600

AGTGCAACTT GCAAACTACT AGTAAGAGGA GAAAAACAAA ATGGTTATGA CTGACCCAAT 6660

CGCAGACTTC CTAACTCGTA TTCGTAATGC TAACCAAGCT AAACACGAAG TACTTGAAGT 6720

ACCTGCATCA AACATCAAAA AAGGGATTGC TGAAATCCTT AAACGCGAAG GTTTTGTAAA 6780

AAACGTTGAA ATCATTGAAG ATGACAAACA AGGCGTCATC CGTGTATTTC TTAAATACGG 6840

ACCAAATGGT GAGAAAGTTA TCACTAACTT GAAACGTGTT TCTAAACCAG GACTTCGTGT 6900

CTACAAAAAA CGTGAAGACC TTCCAAAAGT TCTTAACGGA CTTGGAATTG CCATCCTTTC 6960

AACTTCTGAA GGTTTGCTTA CTGATAAAGA AGCACGCCAA AAGAATGTTG GTGGTGAGGT 7020

TATCGCTTAC GTTTGGTAAA ATCAAGATAC AAAGCTCGTA AAGAACAAAG CAAAATTAGG 7080

AAGTTGGAGA AGTTTGTTTA CAAACAAGCC AACTTATCTA TTTTGCACAG TTCTTAGAGC 7140

GTGTTCAGTT CAGCTCTTGA ACTAAATAAG TATCTGAACC CCGTGAAAAC TGGCCGTTCT 7200

GGCCTGACAA TTTAACAGGA GAAAATAAAC ATGTCACGTA TTGGTAATAA AGTTATCGTG 7260

TTGCCTGCTG GTGTTGAACT CGCTAACAAT GACAACGTTG TAACTGTAAA AGGATCTAAA 7320

GGAGAACTTA CTCGTGAGTT CTCAAAAGAT ATTGAAATCC GTGTGGAAGG TACTGAAATA 7380

ACTCTTCACC GTCCAAACGA TTCAAAAGAA ATGAAAACTA TCCACGGAAC TACTCGTGCC 7440

CTTTTGAACA ACATGGTTGT TGGTGTATCA GAAGGATTCA AGAAAGAACT TGAAATGCGT 7500

GGGGTTGGTT ACCGTGCACA GCTTCAAGGA TCTAAACTTG TTTTGGCTGT TGGTAAATCT 7560

CATCCAGACG AAGTTGAAGC TCCAGAAGGA ATTACTTTTG AACTTCCAAA CCCAACAACA 7620

ATCGTTGTTA GCGGAATTTC AAAAGAAGTA GTTGGTCAAA CAGCTGCTTA CGTACGTAGC 7680

CTTCGTTCAC CAGAACCATA TAAAGGTAAA GGTATCCGTT ACGTTGGTGA ATTCGTTCGC 7740

CGTAAAGAAG GTAAAACAGG TAAATAATGT TGAGTGGTTG ATCATCAACC ACCAACCTAT 7800

TTTCCAACTT TGTGCATAGC ACACGATTTA AAACTAAAGA GGTGAAAACT GTGATTTCAA 7860

AACCAGATAA AAACAAACTC CGCCAAAAAC GCCACCGTCG CGTTCGCGGA AAACTCTCTG 7920

GAACTGCTGA TCGCCCACGT TTGAACGTAT TCCGTTCTAA TACAGGCATC TACGCTCAAG 7980

TGATTGATGA CGTAGCGGGT GTAACGCTCG CAAGTGCTTC AACTCTTGAT AAAGAAGTTT 8040

CAAAAGGAAC TAAAACTGAA CAAGCCGTTG CTGTCGGTAA ACTCGTTGCA GAACGTGCAA 8100

ACGCTAAAGG TATTTCAGAA GTGGTGTTCG ACCGCGGTGG ATATCTATAT CACGGACGTG 8160

TGAAAGCTTT GGCTGATGCA GCTCGTGAAA ACGGATTGAA ATTCTAATAG GAGGACACTA 8220

GAAAATGGCA TTTAAAGACA ATGCAGTTGA ATTAGAAGAA CGCGTAGTTG CTGTCAACCG 8280

TGTTACAAAA GTTGTTAAAG GTGGACGTCG TCTTCGTTTC GCAGCTCTTG TTGTTGTTGG 8340

TGACCACAAT GGTCGCGTAG GATTTGGTAC TGGTAAAGCT CAAGAAGTTC CAGAAGCAAT 8400

CCGTAAAGCA GTAGATGATG CTAAGAAAAA CTTGATCGAA GTTCCTATGG TTGGAACAAC 8460

AATCCCACAC GAAGTTCTTT CAGAATTCGG TGGAGCTAAA GTATTGTTGA AACCTGCTGT 8520

AGAAGGTTCT GGAGTTGCCG CTGGTGGTGC AGTTCGTGCC GTTGTGGAAT TGGCAGGTGT 8580

GGCAGATATT ACATCTAAAT CACTTGGTTC TAACACTCCA ATCAACATTG TTCGTGCAAC 8640

TGTTGAAGGT TTGAAACAAT TGAAACGCGC TGAAGAAATT GCTGCCCTTC GTGGTATTTC 8700

AGTTTCTGAT TTGGCATAAG AAAGGGGATA AAATGGCTCA AATTAAAATT ACTTTGACTA 8760

AGTCTCCAAT CGGACGCATT CCATCACAAC GTAAAACTGT TGTAGCACTT GGACTTGGCA 8820

AATTGAACAG CTCTGTTATT AAAGAAGATA ACGCTGCTAT CCGTGGTATG ATCACAGCAG 8880

TATCTCACTT AGTAACAGTT GAAGAAGTAA ACTAATGAaG TTTTAGGGGA TGTGCACTGT 8940

ACCATCCCCT AAAACTAGAT ATAGTCATCT ATGATGACAT CGTATAGGCG AGTTGATGGG 9000

GGAGACAACC TTTTCTCCCT TATCGGCCCT AGCATTTTAC AAAAGAGGAG AAAATAAAAA 9060

TGAAACTTCA TGAATTGAAA CCTGCAGAAG GTTCTCGTAA AGTACGTAAC CGCGTTGGTC 9120

GTGGTACTTC ATCAGGTAAC GGTAAAACAT CTGGTCGTGG TCAAAAAGGT CAAAAAGCTC 9180

GTAGCGGTGG CGGAGTTCGC CTTGGTTTTG AAGGTGGACA AACTCCATTG TTCCGTCGTC 9240

TTCCAAAACG TGGATTCACT AACATCAACG CTAAAGAATA CGCAATTGTG AACCTTGACC 9300

AATTGAACGT CTTTGAAGAT GGTGCTGAAG TAACTCCAGT TGTTCTTATC GAAGCAGGAA 9360

TTGTTAAAGC TGAAAAGTCA GGTATTAAAA TTCTTGGTAA CGGTGAGTTG ACTAAGAAAT 9420

TGACTGTGAA AGCAGCTAAA TTCTCTAAAT CAGCTGAAGA AGCTATCACT GCTAAAGGTG 9480

GTTCAGTAGA AGTCATCTAA GAGAGGTGAC CTATGTTTTT TAAATTATTA AGAGAAGCTC 9540

TTAAAGTCAA GCAGGTTCGA TCAAAAATTT TATTTACAAT TTTTATCGTT TTGGTCTTTC 9600

GTATCGGAAC TAGCATTACA GTTCCTGGTG TGAATGCCAA TAGCTTGAAT GCTTTAAGTG 9660

GATTATCCTT CTTAAACATG TTGAGCTTGG TGTCGGGGAA TGCCCTAAAA AACTTTTCGA 9720

TTTTTGCCCT AGGAGTTAGT CCCTATATCA CCGCTTCTAT TGTTGTCCAA CTCTTGCAAA 9780

TGGATATTTT ACCCAAGTTT GTAGAGTGGG GTAAACAAGG GGAAGTAGGT CGAAGAAAAT 9840

TGAATCAAGC TACTCGTTAT ATTGCTCTAG TTCTCGCTTT TGTGCAATCT ATCGGGATTA 9900

CAGCTGGTTT TAATACCTTG GCTGGAGCTC AATTGATTAA AACTGCTTTA ACTCCACAAG 9960

TTTTTCTGAC GATTGGTATC ATCTTAACAG CTGGTAGTAT GATTGTCACT TGGTTGGGTG 10020

AGCAAATTAC AGATAAGGGA TACGGAAACG GTGTTTCCAT GATTATCTTT GCCGGGATTG 10080

TTTCCTCAAT TCCAGAGATG ATTCAGGGCA TCTATGTGGA CTACTTTGTG AACGTCCCAA 10140

GTAGCCGTAT CACTTCATCT ATCATTTTCG TAATCATTTT GATTATTACT GTATTGTTGA 10200

TTATTTACTT TACAACTTAT GTTCAACAAG CAGAATACAA AATTCCAATC CAATATACTA 10260

AGGTTGCACA AGGTGCTCCA TCTAGCTCTT ACCTTCCGTT AAAAGTAAAC CCTGCTGGAG 10320

TTATCCCTGT TATCTTTGCC AGTTCGATTA CTGCAGCcTG CGGCTATTCT TCAGTTTTTG 10380

AGTGCCACAG GTCATGATTG GGCTTGGGTA AGGGTAGCAC AAGAGATGTT GGCAACTACT 10440

TCTCCAACTG GTATTGCCAT GTATGCTTTG TTGATTATTC TCTTTACATT CTTCTATACG 10500

TTTGTACAGA TTAATCCTGA AAAAGCAGCA GAGAkCCTAC AAAAGAGTGG TGCCTATATC 10560

CATGGAGTTC GTCCTGGTAA AGGTACAGAA GAATATATGT CTAAACTTCT TCGTCGTCTT 10620

GCAACTGTTG GTTCCCTCTT CCTTGGTGTG ATTTCCATTT TACCGATTGC AGCTAAAGAT 10680

GTATTTGGTC TTTCTGATGT TGTTGCCTTT GGTGGAACAA GTCTCTTGAT CATTATCTCT 10740

ACAGGTATCG AAGGAATCAA GCAATTGGAA GGTTACCTAT TGAAACGTAA GTATGTTGGT 10800

TTCATGGACA GAACAGAATA AAAGTATTTA CTGAATCAGT AAATACTGAG GGAGTGGAGG 10860

TTTAAACTCT GACATTTGTA AGAGTTGGAT CTCCCCTCTT CTATTTTGTT TTTAAATCGG 10920

GGTGAAAAGA CTTTTTGCTT CTATTTAAAA ATAAAATAAG GAGATCAAAT CATGAATCTT 10980

TTGATTATGG GCTTACCTGG TGCAGGTAAG GGAACTCAAG CAGCAAAAAT CGTAGAACAA 11040

TTCCATGTTG CACATATCTC AACAGGTGAT ATGTTCCGCG CTGCAATGGC AAATCAAACT 11100

GAAATGGGTG TTCTTGCTAA GTCATATATT GACAAGGGTG AATTGGTTCC TGACGAAGTT 11160

ACAAATGGAA TCGTAAAAGA ACGCCTTTCA CAAGATGATA TTAAAGAAAC AGGATTCTTA 11220

TTGGATGGTT ACCCACGTAC AATTGAACAA GCTCATGCCT TGGACAAAAC ATTGGCTGAA 11280

CTTGGCATTG AACTAGAAGG TGTTATCAAT ATTGAAGTGA ACCCTGACAG CCTTTTGGAA 11340

CGTTTGAGTG GGCGTATCAT CCACCGCGTA ACTGGAGAAA CTTTCCACAA GGTCTTTAAC 11400

CCACCAGTTG ACTATAAAGA AGAAGATTAC TACCAACGTG AAGATGATAA GCCTGAGACA 11460

GTAAAACGTC GTTTGGATGT TAATATTGCT CAAGGAGAAC CAATCATTGC TCACTACCGT 11520

GCCAAAGGTT TGGTTCATGA CATCGAAGGT AATCAAGATA TCAATGATGT CTTCTCAGAT 11580

ATTGAAAAAG TATTGACAAA TTTGAAATAA AGCGTTTTTC ACACTTGCAA AAATCCGCTA 11640

CAAATGTTAT ACTGAGATAG TCTGACTTAT AATTGTTGTC TCTGTGTCTA GAGGCATCGA 11700

ATCGAAATTT ATGGAGGTGC TTTTGCGTGG CAAAAGACGA TGTGATTGAA GTTGAAGGCA 11760

AAGTAGTTGA TACAATGCCG AATGCAATGT TTACGGTTGA ACTTGAAAAT GGACATCAGA 11820

TTTTAGCAGG G 11831

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10726 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

CCCGGCATTT GAAAGCTATT CGTGAAGGAT TTATGATGGC AATGCCTTTG ATTTTAGTCG 60

GCTCTTTATT TCTTATTCTA ATCAGTTGGC CTCAAGAGGC TTTTACAAAT TGGCTGAATA 120

GTGTTGGATT GCTAAGTATC TTGACAACTA TGAATCAGTC AACAGTAGCG ATTATCTCCT 180

TGGTCGCTTG TTTCGGTATT GCCTACAGGT TGTCGGAAGG ATATGGTACA GATGGTCCGT 240

CGGCAGGGAT CATAGCCTTA TCCAGTTTTG TATTGATGGC ACCTCGTTTT TCGAGTATGG 300

TTTATGATAA AAATGGGGAG CAGGTCAAGC AGTTATTTGG CGGCGCAATA CCATTTTCTA 360

GCCTGAATGC ATCTTCTTTG TTTATGGCGA TTACTATTGG ATTGGTTACA GCAGAGATTT 420

ATCGTATGTT TATCCAGCGC GGAATTACGA TAAAAATGCC AAGTGGTGTC CCAGATGTAG 480

TAAGTAAATC ATTTTCAGCT CTTTTATCTG GTTTTACTAC TTTTGTTTTG TGGGCTTTGG 540

TCTTAAAAGG TCTTGAAGCG GCAGGAGTTG CAGGAGGTCT CAACGGACTC CTAGGTGCAA 600

TTGTTGGAAC ACCGCTTAAG TTAATTGCAG GAACGCTTCC AGGTATGATT CTATGTGTTA 660

TTGTAAACTC ATTCTTTTGG TTCTGTGGAG TTAATGGGGG ACAAGTTTTA AATGCTTTTG 720

TAGACCCAGT TTGGTTACAA TTTACTACAG AAAACCAAGA AGCTGTGGCT GCAGGACAAA 780

CACTCCAACA CATTATTACA TTACCGTTTA AAGATTTATT TGTATTTATT GGTGGCGGTG 840

GAGCGACTAT TGGTCTTGCG ATTTGTCTCT TCCTATTTAG TAAGAGTCGT GCGAATAAAA 900

CATTAGGTAA GCTAGCTATT ATACCGTCTA TTTTTAATAT CAATACAGCT ATTCTATTTA 960

CGTTTCCAAC AGTTTTAAAT CCGATTATGC TGATTCCGTT TATTGCTACT CCTACAATCA 1020

ATGCCTTGAT TACCTATGTA TCAATGGCTG TAGGATTAGT ACCCTATACA ACAGGTGTAA 1080

TCCTTCCGTG GACAATGCCA CCGATTATAG GAGGCTTCCT TGCAACAGGG GCTAGTTGGC 1140

GAGGAGCTCT ATTACAAGTT GTTTTGATTT TGGTTTCTGT AGCAATTTAT TATCCATTCT 1200

TCAAAATTGC AGATAAACGC AATCTTGAAA AAGAAAAAGC TACTGTTGGA GGGAAATAAG 1260

ATGGTTATCA GAGTATTTGA TCAACAGAAA AATACTTATT CTAGCTTTGC CTTAGAGGAA 1320

TTAAGTTACT ATATGAATCG GGTCTTTAAG ACTAACATAG AGCTTGTCGA GGAGAAGGAA 1380

GCGGATATTT TTGTAGGATT AGTCAATAAA GAGGACAGAA AAGACCATGT TCTTATCTCA 1440

TTAGACAAGG GTAAGGGGAG AATTGAGTCT AATACAATTG TAGGTTTACT TATTGGAATT 1500

TACCGAATGT TTCATGAATT TGGGGTTGTG TATACTAGAC CAGGGCGCAG ACATGACTTT 1560

GTTCCAGAGT TACGATTTGA AGATTTTTTA GATAAACAGC TATCTATAGA TGAAACAGCC 1620

AGTTACTATC ATAGGGGAGT ATGTATAGAG GGAGCGGATT CATTTGAAAA TATACTAGAT 1680

TTCATTGATT GGCTACCTAA GATTGGGATG AACAGTTTTT TCATCCAGTT TGAAAATCCT 1740

TACTCTTTTT TGAAACGTTG GTATGAACAT GAATTTAATC CATATCTAAA TAAAGAACAA 1800

TTTTCAAATG AATTAGTACA AGAATTGAGT GATAGGTTGG ATAAAGAATT GCAAAAAAGA 1860

GGTCTTATTC ATCATCGTGT TGGTCATGGA TGGACAGGTG AAGTTTTAGG TTACTCTTCA 1920

AAATTTGGCT GGGAATCAGG TCTTAGTATT TCAGAGGAGA AGAAACCCTA TGTCGCTGAA 1980

ATAAACGGGA AACGAGAATT GTTTAATACG GCTCCGATTT TAACCAGCCT GGATTTTTCA 2040

AATCCAGATG TAGCTGATAA GATGGTAGAA ATTATCAAGG ATTATGCCAA GAAAAGACCT 2100

GATGTTAACT ACTTACATGT ATGGTTGTCG GATGCTCGTA ATAATATTTG TGAATGCGAA 2160

AACTGTAGAC AAGAATTGGT TTCGGATCAG TATATTCGTA TTCTCAATCA ATTGGATAGG 2220

GCTTTAACGA GTGAGGGATT AGATACAAAG ATTTGTTTTC TGCTTTATCA TGAGTTGTTA 2280

TGGGCACCTC AGAAAGAAAA ATTAGATAAT CCTGAACGCT TTACCATGAT GTTTGCACCG 2340

ATTACAAGAA CATTTGAAAT GAGTTATGCA GATGTAGATT TTGACAATTC CATACCTACG 2400

CCTAAACCTT ATATGCGTAA TAAAATTATA CTTCCGAATT CTCTTGAGGA AAATTTATCT 2460

TATCTTTTTG AGTGGCAAAA AGCATTTAAA GGAGATAGTT TCGTATATGA CTATCCTTTA 2520

GGGCGTGCTC ATTATGGCGA TTTAGGCTAT ATGAAAATTA GTCAAACTAT TTACAGAGAT 2580

GTATCTTATC TTTCCAACCT ACATTTGAAC GGGTACATTT CGTGTCAAGA ATTACGTGCC 2640

GGATTCCCTC ATAATTTTCC TAATTATGTC ATGGGGGAAA TGCTCTGGAA GAAGACAAGA 2700

AGTTATGAAG AATTGATTGA AGAATACTTT TCTGCTTTGT ATGGGGAAAA TTGGCAGTCT 2760

GTTGTTGAAT ATTTAGAAAA ATTATCGATT TATTCGTCTT GTGATTATTT TAATGCAATT 2820

GGCAGCCGTC AAAGTGATGT TTTAGCGAAT CATTATTATA TAGCTTACAA TCTAGCTGAT 2880

AATTTTTTAC CAATTATTGA GGAAAATATT TCTAAGTTAT TAAATAGTCA AAAGGATGAA 2940

TGGAAACAGC TCAGTTATCA TCGTGAATAT GTTGTTAAGA TGGCGAAGGC TTTATATCTT 3000

CAAGCAACTG GAAAAACAAG GCAAGCTCAA GATGAATGGA GAAATGTGTT GAATTATATC 3060

CGTGGGCACG AATTGCTATT TCAATCTAAT TTGGATGTTT ATCGTGTAAT TGAAGTAGCA 3120

AAAAATTACG CTGGTTTCCA CTTATAAATC ATAAGTATAG AAAATGAACT AAGGTATTCA 3180

GAGAAGATTG ATCCTAAATA TTATGAAATT TAAGGATTTT TAAGATATTT AGGGTCAACT 3240

TTCTATTTAT ATCGTAGCGA AGTCATTTTA ATAATGATGT GTAAAAGATG GATCAAGATT 3300

GAGGAGGAAG AAAGATGAAA TCAAAAGAAG AAATAAATAT GCTTGGTTTT ACAATTGTCG 3360

CTTACGCAGG AGATGCAAGG TCAGATTTGA TGGATGCTTT GGCGTTTGCG AGAGATGGAT 3420

ATTTTGAACA GGCAAGAGAA TTGGTTGAGT CTGCAAACGA CTCAATAGTG TCTGCCCATC 3480

GAGAACAGAC TAATTTATTA GCGGAGGAGG CATATGGAGA TAATTTTGAA GTGAGCTTTA 3540

TTATGATTCA TGGTCAAGAT ACTTTGATGA CAACGATGCT ATTGTATGAT CAGGTAAAGT 3600

TTTTTATTGA TGAATATGAA CGAATTCGAA AGATTGAAGA ACATATTGGT TTGCAATGAG 3660

GATTAGTCAT GGAAAATTTA CAGGTTAAAG CCTTACCGAA GGAGTTTTTA TTAGGAACTG 3720

CTACCGCTGC TTATCAAGTA GAGGGTGCAA CTAGGGTAGA TGGCAAAGGA ATAAATATGT 3780

GGGATGTTTA TTTGCAAGAA AATAGTCCGT TCTTACCAGA TCCAGCTAGT GATTTTTATT 3840

ATCGTTACGA AGAGGATATA GCTTTGGCGG CAGAACATGG TTTGCAGGCT TTGCGTTTAT 3900

CTATTTCTTG GGTTCGTATA TTTCCTGATA TAGATGGGGA TGCTAATGTA TTAGCTGTTC 3960

ATTATTACCA TAGAGTTTTT CAGTCTTGCT TAAAACATAA TGTGATTCCG TTTGTTTCTT 4020

TACATCATTT TGATTCGCCT CAGAAAATGT TAGAAACAGG GGATTGGTTG AACAGAGAGA 4080

ATATTGATCG TTTCATACGA TATGCTCGCT TTTGTTTCCA AGAATTTACA GAAGTCAAGC 4140

ATTGGTTTAC AATCAATGAA CTGATGTCTC TTGCTGCAGG TCAATATATA GGAGGTCAGT 4200

TTCCTCCAAA TCATCATTTT CAATTATCTG AAGCAATTCA AGCGAATCAT AATATGTTGT 4260

TGGCGCATGC TCTTGCAGTC CTCGAATTTC ATCAATTAGG GATTGAGGGA AAGGTAGGTT 4320

GTATTCATGC TTTAAAGCCA GGCTATCCTA TTGATGGGCA AAAAGAAAAT ATTTTGGCAG 4380

CTAAACGGTA TGATGTTTAT AATAATAAAT TTCTATTAGA TGGAACTTTT TTGGGCTACT 4440

ACAGTGAGGA CACGCTTTTT CACTTGAATC AAATATTGGA AGCTAATAAT TCTAGCTTTA 4500

TTATTGAAGA TGGTGATTTA GAAATTATGA AGAGAGCTGC ACCTCTTAAT ACGATGTTTG 4560

GGATGAATTA TTATCGTTCA GAATTTATTC GTGAATACAA AGGTGAAAAT AGACAAGAAT 4620

TTAATTCAAC AGGAATAAAA GGACAGTCTT CTTTTAAATT AAATGCTCTA GGTGAATTTG 4680

TAAAAAAACC TGGTATTCCG ACAACAGATT GGGATTGGAA TATTTATCCT CAAGGGTTAT 4740

TTGATATGTT GCTTCGTATC AAAGAAGAAT ATCCTCAACA TCCGGTCATT TATTTAACTG 4800

AAAATGGTAC AGCCCTTAAA GAAGTTAAGC CAGAGGGCGA GAATGATATT ATTGATGACA 4860

GTAAGAGAAT CCGTTATATT GAGCAACATT TACACAAAGT TTTAGAGGCT CGAGATAGAG 4920

GAGTCAATAT TCAAGGCTAT TTTATATGGT CTTTGCAAGA TCAATTTTCT TGGGCGAATG 4980

GCTACAATAA GCGATATGGT CTTTTCTTTG TTGATTATGA AACACAGAAG AGATATATTA 5040

AGAAAAGTGC TCTTTGGGTA AAAGGGCTAA AACGGAATTA AGGTTAGCGA TTTGACTGAT 5100

GTTTAATATG TTTTAAATAT GAGGTTGAAT TTTTTATAGG AGGAGTTTTA TGGATAAGCT 5160

AGTCGCTGCC ATTGAAAAGC AACAAGGGAA ATTTGAAAAA ATTTCTACTA ATAACTATAT 5220

GATGGCTATT AAAGATGGAT TCATTGCTAC TATGCCTTTA ATTATGTTTT CAAGCTTTTT 5280

GATGATTATT ATTATGATTC CTAAAAATTT CGGAGTAGAG TTACCGAGTC CAGCTATTGT 5340

CTGGATGAGA AAAGTGTATA TGTTAACCAT GGGAGTTTTG GGTATTATTG TTTCAGGGAC 5400

TGTTGGAAAG TCATTAGTTG GAAATGTTAA CAGAAAAATG CCTCACGGAA AGGTAATAAA 5460

TGATATTTCT GCAATGTTGG CAGCCATATG TAGTTATCTG GTATTAACTG TAACGCTTGT 5520

AGTTGATGAG AAGACGGGAT CTACAAGTTT GTCGACAAAC TATTTAGGAT CTCAAGGATT 5580

GATAACTTCG TTTGTCAGTG CCTTTATTAC TGTAAATGTT TACCGATTCT GTATTAAGCG 5640

AGACATTACT ATTCATTTAC CTAAGGAAGT TCCTGGGGCT ATATCACAAG CTTTTAGAGA 5700

TATTTTCCCT TTTTCTTTTG TTTTACTTAT TAGTGGTTTG TTAGATATTG TATCTCGGTT 5760

TAGTTTAGAT GTTCCTTTTG CCCAAGTATT TCAACAACTA TTGACTCCTA TTTTTAAGGG 5820

GGCAGAATCA TATCCTGCTA TGATGTTGAT TTGGTTTATG TGTGCTTTGC TTTGGTTTGT 5880

TGGAATTCAT GGACCATCTA TTGTCTTACC TGCTGTTACA GCTTTGCAAC TGAGCAATAT 5940

GGAAGAGAAT GCTCAACTTC TTGCAAATGG GCAGTTCCCT TATCATTCTT TAACACCTAA 6000

TTTCGGGAAT TATATCGCTG CTATTGGAGG AACGGGGGCT ACCTTTGTTG TACCATTTAT 6060

TTTGATTTTC TTTATGCGGT CTAAACAATT AAAATCGGTA GGTAAAGCTA CAATTACTCC 6120

TGTTTTATTT GCGGTAAATG AACCTCTTCT ATTTGGTATG CCTGTTATTT TGAATCCCTA 6180

TCTTTTTGTC CCTTTTTTGA TGACTCCACC AGTGAATGTA TTTCTAGGAA AGGTCTTTAT 6240

TGATTTCTTT GGAATGAATG GATTTTATAT CCAGTTACCT TGGACCTTTC CTGGTCCCTT 6300

GGGATTGTTA ATTGGAACGA ATTTTCAACT TATCTCCTTT GTATTTTTAT CTTTGATTTT 6360

AGTTGTCGAC ATATTGATTT ATTTGCCATT CTGTAGAGCG TATGATAGAC AGTTACTGGT 6420

GAAAGAAGAT ATTGCAAGCT CAAATGATAT TATTTTAGAG GAGGATACAA GTGAAATAAT 6480

TCCTGGTGAG ATAGATGAAA TAAAAAGTAA GGAGTTGAAA GTACTGGTTC TTTGTGCAGG 6540

GTCTGGAACA AGTGCGCAAT TAGCCAATGC AATTAACGAG GGGGCTAACT TAACAGAGGT 6600

TAGAGTGATT GCGAATTCAG GAGCGTACGG AGCTCATTAT GATATTATGG GTGTTTATGA 6660

TTTAATTATT CTGGCCCCAC AAGTTCGGAG TTATTATAGA GAGATGAAGG TGGATGCAGA 6720

AAGATTAGGT ATTCAGATAG TTGCTACCAG AGGAATGGAA TATATTCATT TAACAAAGAG 6780

TCCAAGTAAA GCCTTACAAT TTGTATTGGA GCATTACCAA GCTGTGTAGT AAGTTTTTCC 6840

ATCTTTTATT TGAGTAAAGA TTTTGTTTAC AGATAGGCTT GGATTTAAAA ACGTTCCCCC 6900

TTTTTTAATA TAAGAATCCC TCTTTCACAA TTGTAAAAAG AGGGATTTTG TATTTTATCT 6960

CTTAGACCAA GTTCTCTTCA TAAAGAGAAG GAGGATTGGG TAAATCTCCA AGCGCCCTGC 7020

AATCATTGCA AAGGATAGGA GAATTTTTGA GATGGGACTA AAGATTGAGA AACTAGAAGT 7080

GGTTCCTAGA ATAGGCCCGA TATTATTGAA ACAGCTAAAG ACAGCGCTGG TCACGACCAG 7140

AAAATCATTG CTATCTAGGC TGACAATAAA GATAAGCGCT AGCAAAATCA TAGCATAGAT 7200

GACAAAGTAC TTGAGAATCT TATGCTGGGT ATCTTTGTCA ATCACCGTTT TATTAACATG 7260

GAGGGTCAAA ACACGGTGGG GCGATAGGAT TGACAAAATT TGGTTTTTGG CAATTTTTGA 7320

AAGGATGAGG CCTCGAATAA TCTTGAGTCC ACCTGCAGTT GATCCAGCAG AGCCACCGAT 7380

TGCCATGAGG AAAAGGAGGA TAAACTGGGA GAAGAGGGGC CAGTTGGTAA TATCTCCATA 7440

TCCAAAACCA GTTGTTGTAA TGATGTTGGA AACCTGGAAG AAGGTCATTT CAAAGCTCTT 7500

TGAAAACCCT GGGTAGAGGT AGAGGGTGTT GAGGCTAATC AAGCCTGTAG AAACCAGTAC 7560

AATGACCAAG TAAGCCCTAA GCTCTTCATC TCCAAAGAAG GCCTTGATGC GACGGAGCAT 7620

GAGGTAGTAG TAGAGGTTGA AATTTACTCC AAAAACCAGA ACTCCGATAC TGACCAGATA 7680

GGTAATCAGT GAGCTGCCAT AGTGGGCAAT TCCGTCGTTA TAGACGGTAA AGCCTCCAGT 7740

TCCCGCTGTC CCCATAGCAA TAACAAAACT ATCGTAGAGA GGCATACCGG CTAGATAATA 7800

GATGATGACA AAGAGGGAGA AGAGAGCTAG ATAAAGGAGA TAGAGAATCT GGGCAGTGTT 7860

TTTTAGTTTG GATACAACCT TGCCAAAAAC AGGACCTGGA ACCTCAGCCT TCATCACCTC 7920

TAGGTGGCTA TTTTTGGCAT TGTCCATAAT AGCAAGTGCA AAAACAAGCA CTCCCATCCC 7980

TCCAATCAAG TGGGTAAAAC TTCGCCAGAA GAGGAGGGAA CGGCTGAGAA CCGAAACGTC 8040

GTTCAAAATA CTTGCTCCAG TAGTTGTAAA TCCAGAACTA ATTTCAAAAA AGGCATCAAT 8100

AAGGCTGGGG ATTTGCCCAG AAAAGACAAA GGGGAGACCA CCAAAGAAAG ACCAAAGGAT 8160

CCAACAGAGG GCAACGATCA AGACTCCCTC CTTGGCATAA ATCCGTTGAT TTTTTGGCTT 8220

CTGTAAACTC CCTGAACCGC CTAACAATAC GAGAATCCCT ATGGTCGAAA AGAGGGCTGT 8280

AAAGACTTGG CTCGATTCAC GGTAATAGAC AGCAATCGCA ACAGGAACCA AAAGAAGAAC 8340

AGCTTCAATC AAAAGTAATT TTGAAAGGAG GTAACGAATC ATACTTTTAT TCATTTCTTA 8400

CCTCGCGATC AAGTCATAAA TCTTGGTGAT GTTTGGCAAC AAGGTTGTTA CTAGGAGCTT 8460

GTCTCCAACT TCCAACATAT CCTCCCCAGT TGGGAAAATA GTCTTGCCCT TTCGAATAAT 8520

GGCTGCAATA AGAACCCCTT TTTTCAATTT CAGTTGAGAA AGAGGTTTGG CAGTCATTTT 8580

ATTGGCTTCC TTGATATGGA ATTGCAGGGT TTCGATTTGG CCATTGGCTA GATGGTGCAT 8640

AGCTTGAAGG TCTGAATACT GGGCATTAAC TCGACCACGA ATAAAGTGCA TAATCGTATC 8700

TACAGCGATG CTTTTAGGTG TGATGATACT TGAAAAATCA GGCGCATTGA TAATCTCGAG 8760

GAGACTGGTA CGATTGACCT TAGTAATATT TTTCTGTACA CCTACCCTGT CAAGGAACAT 8820

AGATGTAATC AGATTTTCCT CATCGACTCC TGTTAGAGTC GCAACGGCAT CATAGTGTTG 8880

AGCACTTTCT TCCAGCAGGA TATCTTTTGC GGTTCCATCT CCTTGAACGA TGTAGAGATT 8940

TGGGAATTTC TCGCTAAAGA AGCTGGCGAT TTCAGGATTG ATTTCAATGA CTTTTGTATC 9000

GATACGACTA TCTTTGAGAA TACCAAGTAG ATAATAGGCA ATTCTACCTG CCCCAACGAT 9060

GAGAAGGCTC TTCACGGCGC GTGATTTAAA ATAATTATGG AAGAGTATCA TATCGACACG 9120

GTTACCAGTG ACAAAGATTC TATCTTTATC CTGTACAGTC ATGTCACCGC TTGGAATGAT 9180

AATTTGATGA TCCCTCTCTA TCGCACAGAC AATGACATTA CCAAATTTTT TACGAAAATC 9240

AGAAATGGGC ATTTGGCAAA GACCGCTGGT GGACTTGACG ACAAATTCCA TGAGGCTAAC 9300

GCGTCCACCA GCAAAGCGTT CGACAGACAG GGCGTTGGGG AAGTCAATGA TATTCGCGAT 9360

AGCGCGGGCA GCCAAGAGCT CAGGATTAAC GATAAGAGAA AAACCGAGAA TATTCTTTTC 9420

CTTGAAATAA GAGTTAGAAT ATTCAGGGTT CCGCACCCGA ACGATAGTTT CTTTAGCTCC 9480

CATTTTCTTG GCTAGAACTG CTGCAATCAT GTTGACTTCA TCGTGCTCAG TCAGGGCGAT 9540

AAAGATATCA CAATCTTGGA CGCTGGCTTG CTCAAGAATG GCAAAATCGG CCCCGTTACC 9600

AAGGATACCA ATGATATCAA AGCGACTGAC AATATGATTG AGAACAGCTT CGTCTTGCTC 9660

AATCAGCAAA ACATCATGCT TTTCTGCAAC CAAGGAGCGA CAGAGGGCAA AACCAACTTT 9720

TCCCCCTCCG ACAAGGATAA TTTTCATAAT AAAACCTACT TTTTCATGAT GTAACTATCA 9780

TACCCTTTTT CAAGAAAAAA TGCACCTACT AGCTAATAAC AAGAGTTTTT AGTGAAAATT 9840

CGCTATAAGG TAAAACTATA CCCTAACCAA TTGAAATAGC TATTAGCGAC TTTCTCTGAA 9900

ATATGGTATG ATAAAGGATA TACAAGGAGA TAAAATGAAT AATAATTTAC TGGTATTACA 9960

ATCAGACTTT GGTCTGGTTG ATGGTGCGGT ATCGGCTATG ATTGGAGTGG CTTTAGAAGA 10020

GTCTCCAACC TTAAAAATAC ATCACTTGAC GCACGATATC ACGCCTTATA ATATTTTTGA 10080

GGGGAGCTAT CGTCTCTTTC AGACGGTGGA TTACTGGCCT GAGGGAACGA CGTTTGTATC 10140

GGTTGTCGAT CCAGGTGTCG GTTCGAAACG TAAGAGTGTA GTTGCCAAGA CTGCAAAAAA 10200

TCAATACATT GTCACGCCAG ATAATGGGAC GCTTTCCTTT ATCAAGAAAC ACGTTGGCAT 10260

TGTAGCCATT CGTGAGATTT CTGAGGTGGC CAATAGGCGT CAAAACACAG AGCATTCTTA 10320

TACCTTCCAC GGTCGTGATG TCTATGCCTA TACTGGTGCT AAACTGGCCA GTGGTCACAT 10380

TACTTTTGAG GAAGTAGGGC CAGAGCTCAG TGTGGAACAG ATTGTAGAGC TTCCAGTCGT 10440

AGCGACCATC ATAGAAGATC ATCTGGTGAA GGGAGCCATT GATATTCTGG ATGTGCGTTT 10500

CGGTTCGCTT TGGACCTCTA TCACACGGGA AGAATTTTAC AAGCTGGAAC CAGAATTTGG 10560

TGATCGTTTT GAAGTGACCA TCTATCATGC TGATATGCTG GTCTATCAAA ATCAGGTTGT 10620

CTATGGCAAA TCATTTGCAG ATGTGAGAAT TGGGCAACCs ATcTTTACrc TCAGCaTCTt 10680

CGATTAGCTG GGCAATTCGT TCTAGTTGGA TTTCGTCAAT CAAGGT 10726

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7163 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

TTATCTTTAA CGATATCAAT CAAGATCTGG TCAATAAAGG GATTGGGGCT TATCGTGAAG 60

TTGGCATCCA AGCCCATGGA TATGTCTGTG ACGTGACAGA CGAGGACGGT ATCCAAGCCA 120

TGGTCAAGCA AATCGAACAA GAGGTTGGTG TCATTGACAT CCTCGTTAAT AACGCTGGTA 180

TTATCCGCCG AGTTCCAATG TGCGAAATGA GCGCCGCTGA TTTCCGTAAG GTCATCGATA 240

TTGACTTAAA CGCACCATTT ATCGTTTCAA AGGCAGTTAT TCCTTCTATG ATAAAGAAAG 300

GGCATGGAAA GATTATCAAT ATTTGTTCGA TGATGAGCGA ACTGGGACGT GAAACAGTTA 360

GCGCTTATGC TGCTGCTAAA GGGGGCTTGA AAATGTTGAC CCGCAACATT GCGTCTGAAT 420

ACGGTGGAGC CAATATCCAA TGTAACGGAA TTGGACCGGG TTATATTGCC ACTCCTCAAA 480

CAGCACCTCT TCGTGAATTG CAAGAAGATG GTTCTCGCCA CCCATTTGAC CAGTTCATCA 540

TTGCAAAAAC ACCTGCTGCA CGTTGGGGAA ATACTGAAGA TTTGATGGGC CCTGCTGTCT 600

TTCTCGCTAG TGATGCCAGC AATTTTGTCA ATGGCCACAT CCTATATGTA GATGGCGGTA 660

TCTTAGCCTA CATCGGAAAA CAACCTGAGT AAAAATAGAA AGAAGATCTT ATGAAAATCG 720

CATTAATCAA TGAAAATAGT CAAGCTAGCA AGAATCACAT TATTTACGAT AGTCTAAAAG 780

AAGCGACAGA TAAAAAAGGC TACCAATTAT TTAACTATGG TATGCGTGGA GAAGAAGGAG 840

AAAGTCAATT AACTTATGTG CAGAACGGAC TAATGGCTGC CATCCTTTTA AATACAAAGG 900

CAGTTGACTT TGTTGTTACC GGCTGTGGTA CGGGTGTAGG GGCTATGCTT GCTTTAAACA 960

GCTTCCCTGG TGTTGTCTGT GGTCTAGCAG TGGACCCAAC TGACGCTTAC CTTTATTCTC 1020

AAATCAATGG TGGTAACGCC TTGTCTATCC CTTATGCCAA AGGATTTGGC TGGGGGGCAG 1080

AACTGACCCT CAAATTGATG TTTGAACGCT TATTTGCTGA AGAAATGGGC GGTGGCTACC 1140

CAAGAGAACG TGTAATCCCT GAACAACGCA ACGCTCGTAT CTTAAACGAG GTGAAACAAA 1200

TCACCCACAA TGATTTGATG ACCATCCTTA AAATAATCGA CCAAGACTTC CTCAAAGACA 1260

CCATCTCTGG CAAATACTTC CAAGAATACT TCTTTGAAAA CTGCCAAGAT GATGAAGTTG 1320

CTGCTTATTT GAAAGAAGTA TTAGCCAAGT AAAGCTATTC TAAACCAGAA AGGAACTAAT 1380

GGATGACGAA AATATTACTG TTTGGCGAAC CATTAATTCG AATTTCACCA TTAGATGCCA 1440

CCAGTATCGG CGATCATGTT GCCAGTTCGA CTTATTTTGG CGGATCAGAA ATTAACATCG 1500

CTTGTAATTT GCAAGCCCTG GGTATCTCAA CGAAAGTTTT TACCGCACTC CCTGCCAACG 1560

AGATTGGAGA TCGTTTTCTC ACATTCTTGA AACAGCACCA AATCGATACC AGTTCAATCT 1620

GTCGGCTTGG CGATCGAATC GGCCTCTACT ATTTGGAGAA CGGCTTTGGT TGTCGTCAAA 1680

GTGAAGTTTT CTACGATCGT AAGCATACGA GTATCAGCCA GATTCGGCCA AACATGCTAG 1740

ATATGGATTC TCTCTTTCAG GGGATTAGCC ATTTTCATTT TAGTGGAATC ACCGTAGCTA 1800

TCGGTCAAGA GGTCCGTGCG ATCCTTCTCC TACTCTTGGA AGAAGCCAAG CGCCGAGGAA 1860

TTGTCGTTTC AATGGATCTC AATCTGAGAA CAAAGATGAT TTCAGTCCTA GAAGCCAAGT 1920

ATGAATTTTC TAAGTTTGCA CGTTTTACTG ACTATTGCTT CGGTATTGAT CCTCTCATGA 1980

TTGATGACCA AAATCTAGAG ATGTTTCCAA GAGACAGTGC TAGCCTAGAA GAGGTGGAAA 2040

ATCGCATGCG ACTTTTAAAA GAAGCCTATG GTTTCAAGGC CATTTTCCAT ACCCTCCGCT 2100

CTAGTGATGA GCAAGACAAA AATGTCTATC AAGCCTATGC TCTAGAAGAA CTATTTGAAG 2160

AGTCTGTCCA ACTAAAAACT GCAGTCTATC AACGAATTGG TAGCGGGGAT GCCTTTATAT 2220

CTGGTGCCCT TTACCAACTA CTCCATCATT CCTCCCTAAA AACTACCATT GACTTTGCAG 2280

TTGCGAGCGC AACTCTCAAA TGCACTCTTC CAGGAGACCA TCTCTCCACT TCCTCAACTA 2340

GTATTGAAAA TTTACTGGCA AATGCACAAG ATATCATTCG TTAGGAGAAT TACATGACCA 2400

AATCAGATAC GATTATTGAA CTAAAAAAAC AAAAAATTGT CGCTGTTATT CGAGGAAATA 2460

CAAAGGAAGA AGGACTACAA GCCTCGATTG CTTGTATCAA GGGCGGTATC AAAGCTATTG 2520

AAATCGCCTA TACCAATCAG TATGCAGGAC AAATCATCAA GGAACTTGTA GACTTGTATC 2580

AGGACGATCA GAGTGTTTGT ATCGGTGCAG GTACTGTGCT TGATGCCGTA ACTGCTAGAG 2640

ATGCCATTCT AGCTGGAGCA AATTACGTTG TTTCTCCATC TTTCCATGCT GAAACTGCGA 2700

AAATGTGCAA TCTCTACAGC ACACCGTACA TTCCAGGCTG TATTACCCTC ACAGAGATCA 2760

CGACTGCACT TGAAGCCGGT AGTGAAATCA TCAAACTCTT CCCAGGTAGT ACTCTCAGTC 2820

CAGCATATAT CTCTGCAGTC AAGGCACCGA TCCCACAAGT TTCCGTAATG GTAACCGGAG 2880

GAGTCGGCCT AAACAACATC CCTCAATGGT TCGCTGCTGG TGCAGATGCC GTTGGAATTG 2940

GTGGCGAACT CAATAAACTC GCTTCCCAAG GCAACTTTGA CCGCATCAGC GAGATTGCCC 3000

AACAGTATAT TACACTCAGA TAAAATCATA ACTACCCGTC TAACGGGTGG TTTATCTCAG 3060

AGCTATAAGC CCAAATCATC AGCCAGCGCC TAAAGACGCT GGCTTTCACG TTGTTCAAGC 3120

CTTATTGCTC TTGACTCGTC ACTTGCCTCT TTAAGAGACT TTGGTATTAC TTACCACTAT 3180

CCCTAAAGGG ATCCTCATAT TCTTTTACAC TCAATTTATC TAGTGCTATA GTAGATTGAA 3240

ACTGGAATAG TACACCTCTG CTTCTAAAAC ATTGTTAAAA ATCGATTTGA CTGTCCTGAT 3300

CGATTTTGTC CTGTTCTTAT TTCATTTTAC TATATATCAT ACTTTACTCG TTCTCAAATT 3360

TTCATACTCA TGAAGAAATC ATCCACTCGA TAATTTCTTT AATCTTGACT ATATTTCTTA 3420

ATTGTGGCTT CATTAAGCCC TACTGGACTT ACATAATAAC CTTCCTCCCA GAAATGCCGA 3480

TTCCCAAACT TGTACTTGAG ATTGGCGTGT TTGTCAAACA TCATGAGTGC ACTTTTGCCT 3540

TTTAAATACC CCATAAAACT TGAAACACTT AGCCTCGACG GAATACTGAC TAACATGTGT 3600

ACATGGTCTG GCATTAAGTG ACCCTCGATC ATTTCAACAC CTTTATAACT ACACAAGCGA 3660

TGAAATATTT CGTCTAAACT ACTTCTATAT TGATTATAGA TGACTTTTCG TCTATACTTA 3720

GGGGTGAACA CAATATGATA GAACACCTCC ACTTTGTGTA TGATAAACTA TGAGTCTTTT 3780

GTGCCATATT TTTTCTCCTT TCGCTTTACA ATTGGATTGA ACACCTTTAT TGTATCGCGT 3840

TTGGAGTTTT TTTGGTATAA CCTTCGACGC GCACCCGTAT AGCGGGTGGT TGTTTTGTCT 3900

CGCACCTCAC GGAGCGAGAC GGACTAATAT AGTGGAGTGA AATAGGATAC GAACAAATTG 3960

ATTAGGAAAA TCAAATGAAT TTATAGAAAT CTTTTAGCAG TTATAACGTT CTATTCTAGT 4020

TTCAAAACGC TATAGTCACA TAATAATGAA GTAAAAAAGG ATAAGTATCA ACTTATCCTT 4080

TTTTAAAAGA AAAATCCGAA GATATTTGGC CTTCTTCGGA TTTTTTCTAT TTTCCACAGT 4140

TTCATGTAAT TCATCTAGAT GATGAACAAA TTAGTTGTTC TTTCCTCTAC GGAATAGATA 4200

AAATGCCCCA AGTAGCAAGA ACCCTAGACT TGCCAAGATT GACTGACCTT CTCCTGTCTG 4260

AGGGAGATTC TTTTGATCCG AATGGTTCTT TTCCTCTTCA GATTTTTCCT TTTCTTTTGA 4320

ATTCTGTACT TGTGGCTGAG CTGCTTGCTC TAGCTTTTTA AAGACTTCCT GATCTGGAGC 4380

TGATTCCTGG GTTTCAGGAT TATAGTAGGC AATCTTATAT TCATCCCCTT CTTTTCGAAT 4440

GGTATAGACT CCACGTTTCA AAACTTGGAA TTGGTTGGAA ATAGTAGAGA CAGAATCATC 4500

ATATTTCACA ATGCCCCAAA CTCCTTGTTT AGCATCATAA ACAGACTGAA GGGTTTCGTT 4560

ATTTTCGATG AGGCTACTTT CTAACTCTTT TATCATTTGA TTGAAGGTGG CACGATCCAC 4620

GTTAGGAATG AGCATATAGC CATAAGAATC TCTATTTTGC TTATGAGCCT GACTAATCGT 4680

AAGAAATTCA TTTTCAACTT CCTTGTCTGA CTGTCCTTCA TTGATATCCT TCCAGGCTCC 4740

CTTTTGCAAA GCCTTACTCA TACTGATTGA ACTCTTCTTA AAGAAAAAGT AACCAATATT 4800

CTTTTTCGAA TCGAACGATT CTAAAAAGAC ACTTTGGGTT TCAGGATAAT CCTTTTCTTG 4860

TTCTGTAAGG GAGGCTTCTT TATCATTGAC ATAGACTTTA TATGGATTAC CTGATTCCAG 4920

TTTTCTCTGG TCAATTGTAG TTGCAGCAGT ATCTGTTGAA GTGTTTTGGA TATTGCTTCC 4980

TAAAAAGGCG ATCTTATCCT TTAGCATAAA CCAGCTCTTA TGAGCAGTCA ATGTTTGATT 5040

CCAGTTGGTG AAATCCATGG TTGCTGTCGC ATTGGCATCA TCTAGTTTGC TCGTTCCAAC 5100

GAAAGCAGAC GGTAAAACTT TACCTGTATC GCTATCCGCT CTCTTAGCAT CCGTCTCTGT 5160

TGTACCAGGC ATCTTATATG GATTAACTGT TGGCCAGTAG CCATCGCTAT AGTGACTCAA 5220

ATCGCCATTG TAAAGATAGA ACATCCCATC ACTCGTATAC CAACCACGTT TATTTTCCTT 5280

GTTCATGTGT TCGTAATTCA AGGTACGACT GGAAAAGAGT GACAAGCCAA ATCCAAACCC 5340

TTTCTCTGCA TTGTACATGG CTGTTTTATC CATCTTGTTA AAGGCAGATA GGTAACTTGG 5400

TCTTGGAACA CTTGCGACTC CTGCATCACT TAACAAGGAT TGCATCAAAC TGATATCCTT 5460

ATAAGTCTTC AAATTCTTAA AGACATCATA ATAACTATCC GATTGAACAA TGGTCTTCAC 5520

AAGACTCTGC AAACATTGTT TGGTTTCTCC TTCAGACATA TCCGCTATTC GGTGAATCCC 5580

TCTTAGTACT TCTACTGCGG CCACGTGCCC CTCGCTATTT GCACGACTGA TCGAGCGTCC 5640

ACGACTCATA TCCATCAACT CTCCATTCAC CAGCAAAGGA GCAAACGATT TATCAATCCA 5700

GTGGTACATG GTTTGCATTT TATCTTTATC GATTGGATTC TTGGTCTTTT GAATGACTGG 5760

CAACAGTTGA GACAGGCCAT CAATCAAAAC ATTCCCATAA GCACCCGTAT AGGCAACATT 5820

GGTGTGGTCG ATATAGGATC CATCTTGATA AAAACCTTCA CCTTGGTCTA CCAACTTGAA 5880

CACTTGCTCA ATCGAGCGAA TGGTAGAAGA AATTTCTTGA TCATCCTTAC GCAGTAAACC 5940

AGCTATTACT TTTACCCTTC CCATATCAAC TAAGTTTCCA CCTAGAGCCT TGAATGGGTT 6000

ATCAGTCGTC TTTCGGAAAT GTTCGGGATC TGGTACAAAT TTTTCAATCA CATCTGTATA 6060

TTTTTTAATT TCCTCATCAG AGAAGTATTC TTTCATCAGA GACAAGGTAT TGTTGATGGC 6120

ACGAGGTGTA CCGATTTCAT AATCCCACCA GTTCCCAACA ATGCTCTTTT CACTATTGTA 6180

GACATGTTTA TGCATCCATT CCATGGAATC CCTGACTGTT CGAACGACAG TTTCATCTTG 6240

ATAATAACGA GAAGAAGGAT TGGTCACTTG CTTGGCCATC TCCTCCAATT TCCGATAAGT 6300

GGCAGTCAGA TTTGCAGACG TTTTATAATT TGAAAATTTT TCCCACAAAT AGGTGCGGTC 6360

CGCCTGACTT GAAATACTGG ATAGGCTATC AGCTACCTTT CCTTCCAATT CCTGGTTTAA 6420

TTTGGCCATC TGTTCATTTT TAGAATCATA GTATTGATTC CCAGCGATGA TGCCATTCCA 6480

GTCATCCAAA CGGTCTGTGT ATGCATCCTT AACAGAGGCC AGAATCTTCA AAGGAATCTT 6540

TTTCACTTCC TTGCCATCTT TACTGACAAT GACATTGGTT GTCCCTTCCT TAAGAGGTTC 6600

TAAAATTCCA TTTTTGACTG AAGCAACGTC AGGATTTTCT ACCTTATAAG TATAGTCCGC 6660

AAGAGAAAAA ACATGTTTTT TTCCAATTGG TAAATCAATC TTTTCCTCAA GCTGTTTATC 6720

TGTTTGAGAA TCCTCAGAAA GCTGGTCTGC TACCTCTACC AGCTCAATAT CCTTAAAGGA 6780

AACAGTCCCA GTTCCTGTTT CATAGAATAA CTCCAGCTTG ATTTTATCAA CATCTAAAGT 6840

CGGGCTATAG TCTGCTTCAA TGGTCTGCCA GTCCTTTGTT CCTGACGTCG TTGCAGAATT 6900

CCACAATCGC TTGTCCTTAC CACTTTCCTC AATGATACGA ACTTTGGCAA TCCCGATTTT 6960

ATTATCTGTT TTAATCTTGA AACGCAGTTT ATACTTTTTC TTAGCTTCAA TAGGAACCAT 7020

ACGGTGAAGC GCTGCCCTTA ATTTCTCATG GCTTGAGATA GTGATAGCCC CATCCTTAGC 7080

CTCAATGACT CGAGTTGAGG CATCTGCACT ATTCTTCTGG TCTACCCAAG CTGACCACCC 7140

CCTGAGCTTT GCTTCCTGTC CGG 7163

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9244 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

CGTTATAACA TACATGTAAG CGGTACCCAA AATGGTGCCA AGTCAAAATT TTTAAGGAGG 60

AAAATACATG TCTTCACATC CAATTCAGGT CTTCTCAGAA ATTGGGAAAC TGAAAAAAGT 120

TATGTTGCAC CGTCCAGGCA AGGAGTTAGA AAACTTGTTG CCGGACTATC TTGAAAGGCT 180

TCTTTTTGAT GATATTCCTT TCTTGGAAGA TGCTCAAAAA GAACATGATG CATTTGCCCA 240

AGCTCTTCGC GATGAAGGAA TTGAGGTTCT CTACCTAGAA CAACTCGCTG CTGAATCATT 300

GACCTCTCCA GAAATCCGCG ATCAATTTAT CGAGGAATAC TTAGACGAAG CCAACATCCG 360

TGATCGTCAA ACCAAGGTTG CTATTCGTGA ATTGCTTCAC GGCATCAAGG ACAACCAAGA 420

ATTGGTTGAA AAAACAATGG CTGGGATTCA AAAAGTTGAA TTGCCAGAAA TTCCTGACGA 480

AGCTAAAGAT CTAACTGACT TAGTTGAATC AGAGTATCCA TTTGCAATTG ACCCGATGCC 540

AAACCTCTAT TTCACTCGCG ACCCATTTGC AACAATTGGA AACGCCGTAT CGCTTAACCA 600

CATGTTTGCA GACACTCGTA ACCGTGAAAC ACTCTACGGT AAGTATATCT TCAAATACCA 660

CCCAATCTAT GGCGGAAAAG TGGATTTGGT CTACAACCGT GAAGAAGATA CGCGTATCGA 720

AGGTGGAGAC GAGTTAGTTC TTTCTAAAGA CGTCCTTGCA GTAGGTATCT CTCAACGTAC 780

AGACGCAGCT TCTATCGAAA AACTTTTGGT CAACATCTTC AAGAAAAATG TTGGCTTCAA 840

GAAAGTTTTG GCCTTTGAAT TTGCTAACAA CCGTAAATTC ATGCACTTGG ATACTGTCTT 900

CACTATGGTA GACTATGACA AGTTCACTAT TCACCCAGAA ATCGAAGGCG ACCTTCACGT 960

TTACTCAGTT ACTTACGAAA ACGAAAAACT TAAAATCGTT GAAGAGAAAG GTGACTTAGC 1020

TGAACTTCTT GCTCAAAACC TTGGTGTAGA AAAAGTTCAT TTGATTCGTT GCGGTGGTGG 1080

CAATATCGTA GCAGCTGCGC GTGAACAATG GAACGACGGT TCTAACACTT TGACCATCGC 1140

ACCTGGTGTG GTAGTTGTTT ATGACCGCAA TACCGTGACC AATAAGATTT TGGAAGAATA 1200

CGGGCTTCGC TTGATTAAGA TTCGCGGAAG TGAATTGGTT CGGGGCCGTG GTGGACCTCG 1260

TTGTATGTCT ATGCCATTTG AACGTGAAGA AGTGTAATCG CTGTTCGATA TTCGTCAATA 1320

GAAAATGTAA AAAATAGAAA GAGGAAATAA TAAAATGACA AATTCAGTAT TCCAAGGACG 1380

CAGCTTCTTA GCAGAAAAAG ACTTTACCCG TGCAGAGTTA GAATACCTTA TTGGTCTTTC 1440

AGCTCACTTG AAAGATTTGA AAAAACGCAA TATTCAACAC CACTACCTTG CTGGCAAGAA 1500

TATCGCTCTC CTATTTGAAA AAACATCTAC TCGTACTCGT GCAGCCTTTA CAACTGCGGC 1560

TATCGACCTT GGTGCTCACC CAGAATACCT CGGAGCAAAT GATATTCAGT TGGGTAAAAA 1620

AGAATCTACT GAAGATACTG CTAAAGTATT GGGACGTATG TTTGACGGGA TTGAATTCCG 1680

CGGATTCAGC CAACGTATGG TTGAAGAATT GGCAGAATTC TCAGGCGTTC CAGTATGGAA 1740

CGGTCTAACT GACGAATGGC ACCCAACTCA AATGCTCGCT GACTACTTGA CTGTTCAAGA 1800

AAACTTCGGT CGCTTGGAAG GCTTGACATT GGTATACTGT GGTGATGGAC GTAACAACGT 1860

TGCCAACAGC TTGCTCGTAA CAGGTGCTAT CCTTGGTGTC AATGTTCACA TCTTCTCACC 1920

AAAAGAACTC TTCCCAGAAA AAGAAATCGT TGAATTGGCA GAAGGATTTG CTAAAGAAAG 1980

TGGCGCACAT GTTCTCATCA CTGAAGATGC TGATGAAGCA GTTAAAGATG CAGACGTTCT 2040

TTACACAGAC GTTTGGGTAT CAATGGGTGA AGAAGACAAA TTCGCAGAAC GTGTAGCTCT 2100

TCTTAAACCT TACCAAGTCA ATATGGACTT AGTTAAAAAA GCAGGCAATG AAAACTTGAT 2160

CTTCCTACAC TGCTTGCCAG CATTCCACGA TACTCACACT GTTTATGGTA AAGACGTTGC 2220

TGAAAAATTT GGTGTAGAAG AAATGGAAGT AACAGACGAA GTCTTCCGCA GCAAGTACGC 2280

TCGCCACTTC GATCAAGCAG AAAACCGTAT GCACACTATC AAAGCTGTTA TGGCTGCTAC 2340

ACTTGGTAAC CTTTATATTC CTAAAGTATA ATTTTAGATA ATAAACCGTC TACCAACAGC 2400

TATGAGGGCT GCGACTAATA GCTTTAGTCC GGTCCTCTTT TATGTAATGG TAATCTATTA 2460

TTTCTTATAA AATATGTGAA AAATCATTAA ATTGAAATCT AAACGCATTC TATTGAGTGT 2520

GATAAAGGAG AATTTATGGC AAATCGTAAA ATTGTAGTAG CTTTGGGAGG AAATGCGATT 2580

CTTTCTTCTG ACCCATCAGC AAAGGCTCAA CAAGAAGCTT TAGTTGAAAC AGCTAAGCAT 2640

CTTGTAAAAT TGATTAAAAA TGGAGATGAT CTGATTATCA CTCACGGTAA TGGACCTCAA 2700

GTTGGGAATC TCTTGCTCCA ACATTTGGCA TCAGACTCTG AAAAGAACCC TGCCTTCCCA 2760

CTCGACTCAC TTGTCGCTAT GACAGAAGGT AGCATCGGTT TCTGGTTGAA AAATGCTTTG 2820

CAAAATGCTC TCTTGGATGA AGGCATCGAA AAAAATGTTG CCTCTGTTGT AACGCAAGTT 2880

GTCGTAGATA AAAATGATCC AGCTTTTGTT AACTTGAGTA AACCAATCGG TCCTTTCTAT 2940

TCAGAAGAAG AAGCAAAAGC AGAAGCCGAA AAAAGCGGAG CGACTTTCAA GGAAGATGCT 3000

GGCCGTGGCT GGCGTAAGGT CGTTGCCTCA CCAAAACCTG TTGACATCAA AGAAATTGAA 3060

ACCATCCGTA CTCTTTTAAA TAATGGTCAA GTCGTCGTAG CTGCAGGTGG TGGCGGTATT 3120

CCCGTCGTCA AAGAAAACAA TGGACATTTG ACTGGTGTCG AAGCGGTTAT TGATAAAGAC 3180

TTCGCTTCCC AACGTTTGGC AGAATTGGTT GATGCAGACC TCTTCATCGT TTTGACAGGT 3240

GTAGATTATG TATTTGTTAA CTACAACAAG CCAAACCAGG AAAAATTGGA ACATGTGAAT 3300

GTTGCCCAGC TGGAAGAATA TATCAAACAA GATCAGTTTG CACCAGGTAG CATGCTTCCA 3360

AAAGTAGAAG CAGCTATCGC TTTTGTCAAT GGTCGTCCAG AAGGAAAAGC AGTTATTACT 3420

TCCCTTGAAA ATCTAGGCGC CTTGATTGAA TCTGAAAGCG GAACAATTAT TGAAAAAGGA 3480

TAAGTTGTTT TACTAATAAG ATGTATTCTA TTTCTAGTAT CTTTATATCA AATTAGAAAT 3540

TATTCTTGAA AACATGTACA ATATTTCAAA AGATACTAGT TTTAGACTTT AATATGGTAA 3600

AACAAATATA AATAGAAAGC GTTTTCTTGA ATGTTTATTT AAGAAAGTAG TTGGTTTTTT 3660

ACACTTTGTT AGACATCAGG AGGAAAAACA AATGAGTGAA AAAGCTAAAA AAGGGTTTAA 3720

GATGCCTTCA TCTTACACCG TATTATTGAT AATCATTGCT ATTATGGCAG TGCTAACTTG 3780

GTTTATCCCT GCGGGGGCCT TTATAGAAGG TATTTACGAG ACTCAGCCTC AAAATCCACA 3840

AGGGATTTGG GATGTCCTCA TGGCACCGAT TCGGGCTATG CTAGGTACTC ATCCAGAGGA 3900

AGGTTCGCTC ATTAAAGAAA CGAGCGCAGC GATTGATGTA GCCTTCTTCA TCCTTATGGT 3960

TGGTGGTTTC CTTGGCATTG TCAACAAAAC TGGTGCTCTT GACGTAGGGA TTGCCTCTAT 4020

CGTGAAGAAG TATAAGGGCC GCGAAAAAAT GTTAATTTTG GTACTGATGC CTTTGTTTGC 4080

CCTCGGTGGT ACAACTTATG GTATGGGTGA AGAAACAATG GCCTTCTATC CACTCCTTGT 4140

GCCAGTTATG ATGGCCGTTG GTTTTGATAG CCTGACTGGT GTTGCAATTA TTTTGCTCGG 4200

TTCTCAAATC GGCTGTTTGG CATCTACTCT GAATCCATTT GCGACAGGTA TTGCTTCAGC 4260

GACTGCGGGA GTTGGTACAG GGGACGGTAT CGTACTTCGT CTGATCTTCT GGGTTACCTT 4320

GACTGCTCTT AGTACTTGGT TTGTTTACCG TTATGCGGAT AAGATTCAAA AAGATCCGAC 4380

TAAGTCACTG GTTTATAGTA CTCGCAAAGA AGATTTGAAA CACTTTAACG TAGAAGAATC 4440

TTCATCTGTA GAATCTACAC TTAGCAGCAA ACAAAAATCA GTTCTCTTCT TATTTGTGTT 4500

GACATTCATC TTGATGGTAT TGAGCTTCAT TCCATGGACA GACCTTGGCG TTACCATTTT 4560

TGATGACTTT AATACTTGGT TGACTGGTCT TCCAGTTATT GGTAATATTG TCGGTTCATC 4620

TACTTCTGCA CTAGGTACTT GGTACTTCCC AGAAGGCGCA ATGCTCTTTG CCTTTATGGG 4680

TATCCTGATT GGTGTTATTT ATGGTCTTAA AGAAGATAAG ATTATCTCTT CCTTCATGAA 4740

TGGTGCTGCT GACTTGCTCA GTGTTGCCTT GATCGTAGCG ATTGCTCGTG GTATTCAAGT 4800

TATCATGAAC GACGGTATGA TTACCGATAC AATCCTCAAC TGGGGTAAAG AAGGCTTGAG 4860

CGGTCTATCT TCACAAGTCT TTATCGTTGT AACTTATATC TTCTATCTAC CTATGTCATT 4920

CTTGATCCCA TCTTCATCTG GTCTTGCCAG CGCAACTATG GGTATCATGG CTCCACTTGG 4980

AGAATTTGTA AATGTCCGTC CTAGCTTGAT TATCACTGCT TACCAATCTG CTTCAGGTGT 5040

CTTGAACTTG ATTGCACCAA CATCTGGTAT TGTGATGGGA GCTCTTGCAC TTGGACGTAT 5100

CAACATTGGT ACTTGGTGGA AATTCATGGG CAAACTCGTA GTCGCTATTA TTGTAGTGAC 5160

CATCGCCCTT CTTCTCCTTG GAACCTTCCT TCCATTCCTA TAAAATAGTG AGTGAGGTGA 5220

TTCCATGAAA ATAGATATAA CAAATCAAGT TAAAGATGAA TTTCTTATAT CATTAAAAAC 5280

CTTGATTTCC TATCCTTCAG TACTCAATGA AGGAGAAAAT GGAACACCTT TTGGACAAGC 5340

AATCCAAGAT GTCCTAGAAA AAACTTTAGA GATTTGTCGA GACATAGGTT TCACTACCTA 5400

TCTTGACCCT AAAGGTTATT ACGGATATGC AGAAATCGGT CAGGGAGCAG AGCTTCTGGC 5460

CATTCTCTGT CATTTGGATG TTGTTCCATC AGGTGATGAA GCAGATTGGC AGACACCGCC 5520

ATTTGAAGCA ACTATCAAAG ACGGCTGGGT ATTCGGACGT GGTGTCCAAG ATGATAAAGG 5580

CCCTTCGCTC GCAGCTCTCT ATGCAGTAAA AAGCTTGCTG GACCAAGGTA TTCAGTTCAA 5640

AAAGCGCGTA CGCTTTATCT TTGGTACCGA TGAGGAAACC CTCTGGCGCT GCATGGCACG 5700

CTACAATACC ATCGAAGAAC AGGCCAGTAT GGGCTTTGCA CCTGACTCAT CTTTTCCTCT 5760

GACCTATGCT GAAAAAGGGC TTCTACAGGT CAAACTTCAT GGCCCTGGAT CGGATCAACT 5820

AGAGCTTGAA GTAGGAGGCG CCTTTAACGT TGTACCAGAC AAGGCCAACT ACCAAGGTCT 5880

CCTCTATGAA CAGGTTTGTA ACGGTCTCAA AGAAGCTGGT TATGATTACC AAACCACTGA 5940

ACAAACCGTA ACGGTTCTCG GAGTGCCAAA GCATGCTAAG GATGCTAGTC AAGGTATCAA 6000

TGCTGTCATC CGACTAGCTA CCATTCTTGC TCCTCTCCAA GAACACCCTG CTCTCAGTTT 6060

TCTTGCAACA CAAGCAGGTC AAGACGGCAC AGGAAGACAA ATCTTTGGTG ATATAGCAGA 6120

TGAACCTTCT GGTCACCTAT CCTTTAATGT CGCAGGTCTC ATGATCAATC ATGAACGTTC 6180

TGAAATCCGT ATTGACATTC GGACTCCTGT CTTAGCTGAC AAGGAAGAAC TAGTAGAGTT 6240

GCTTACAAGA TGTGCACAAA ACTACCAACT CCGCTACGAA GAGTTTGACT ATCTAGCGCC 6300

TCTATACGTC GCAGAAGACA GTAAACTCGT TAGCACACTG ATGCAAATCT ACCAAGAAAA 6360

GACTGGCGAT AACAGTCCTG CTATTTCATC CGGTGGTGCC ACTTTTGCTC GCACCATGCC 6420

AAATTGTGTA GCCTTCGGCG CCTTATTCCC AGGAGCGAAG CAGACAGAAC ATCAGGCAAA 6480

TGAATGTGCC GTTCTAGAAG ATTTGTACCG TGCTATGGAT ATTTATGCCG AAGCCGTCTA 6540

TCGACTTGCA ACTTAATCAG GCAACTGTTT CTACCAAAAA AAATCGACCG ATTAATGAAC 6600

TGCACCCCAA AAGTTAGACA GAATAAATCT AACTTTTGGG GTGTTTTATT ATGAAATTGA 6660

GTTATGAAGA TAAAGTTCAG ATCTATGAAC TAAGAAAGCA AGGACAAAGC TTCAAACAGC 6720

TTTCAAAAAG ATTTGGTGTG GATGTTTCTG GTCTAAAGTC ATCTGAATCT TTGAGATGAG 6780

CTTTATAAAT CGCTTTTTTC AGTTTTTGCA CTGGTGTTTC GATAAACTCA AACTTTTTAG 6840

CCGTGGTATT GCCTGATTTT ATAGTATATT GAAACTAGAA TAGTACACCT CTCCTTCTAA 6900

AACATTTTTA GAAATCGATT TGACTGTCCT GATCGATTTG TCCTGTTCTT ATTTCATTTT 6960

ACTATATTTG AGCCACTTCG TCTTTAACGG CTTTATTCAT AAGCTCTTGT AATTTTTCTT 7020

TACTATCAAT TACTTCTGAT TTTCCGTTGT AATTTATTGT AATAGGTTTT AACTTACCTA 7080

ATTTCTCGAC ACGCTCATTA ATTTGATCTT TTTTGAAGGC TGCTTATGTT TTTCCTAAGA 7140

TTTTTTCAAA AATATATTTA TCAGATAGCG GTTTGTCTTC TTCTTCAGCT TGGTTTTTGT 7200

ATTAATTTGA AACATAAGGA ACAAATCCTT CATAGTAACC TAATGCTCCC ATAAGTTCAA 7260

AAGCTTGTTT TCTAATTCAA ACCATTGCAA CTCAGATTTC AGCTTTTCAG ATAAATCCTG 7320

CTCATCCAAA TAATGACTTG AAATTAGTGC TGAACTCGTT TCTGTATCCT GTACAGGCTG 7380

AGCACCCATA CCAGCAAAAA ATAAACTCGT TCCTAGCAAG ACCGAACAAG CTCCTATTGC 7440

ATATGGCCTC AAAGAAAAAC GCTGCTTTCT CTCAAATTGA AATTCTTTCA TCCCATCTCC 7500

CATCATTCAT TATTACTGTA TATTTTGTAT ATCAGAAATA GTTTGTATTC ACAAATCTTT 7560

CTAGTTATTC CCTTATCATT CCTAATTAAG GGAGATAACA TACAATAATT TTTAGTTAAA 7620

TGTATATCGA TGTTTTTTGT TTTTCTTAAT AAACGCAATA CAAAAAGAGC CTGTTACCAA 7680

GCTCTTTGTA CTCAATGAAA ATCAAAGAGC AAATTAGGAA ACTAGCCACA GGTTGCTCAA 7740

AACACCGTTT TGAGGTTGCA GATAGAACTG ACGAAgTCAG CTCAAAACAC TGTTTTGAGG 7800

TTGCAGATAG AACTGACGAA GTCAGTAACA TCTATACGGC AAGGCGACGC TGACGTGGTT 7860

TGAAGAGATT TTCGAAGAGT ATTAGTCTAT TATTTCTTCT CAGCGCGAAG GGCTGACAAG 7920

ATTTGTGTTC GGATATCATC CACACCATTT GGAGTATTTG GTAAAAAGAT AGTTTGATTT 7980

CCTTTAGAGG CAAAGGTATT CAAGGTATCC AAATACTGGT TGGTCAAGAG GATAGACATG 8040

ATTTGTTCTT CTGTCATGCC AACATTGGCT TCCTTGAGTT CGGTGATAGA CTCTGCCAAT 8100

CCATCCACAA TCGCCTTACG TTGTTGGGCA ATCCCCACAC CATGAAGGCG GTCTTTTTCT 8160

GCTTCTGCTT CAGCTGCAGT GACAATTTTA ATCTTGTCAG CTTCCGCCAA TTCTTGTGCT 8220

GCGACCCGCT TACGTTGCGC CGCATTGATT TCATTCATGG ATTGCTTAAC TTCTGCATCT 8280

GGTTCGACCT TGGTAATCAA GGTTTTCACG ATAATGTAGC CGTAAGTGGT CATTTCTTCT 8340

GCTACTTGGT GTTGAACTTC AAGGGCAATC TCATCTTTTT TCTCAAACAA TTCATCCAAG 8400

GTTAATTTTG GAACAGAAGA GCGAAGAGCA TCTTCGATAT AAGATTTAAT CTGAGATTCT 8460

GGACGTATGA GTTTATAGTA AGCATCTGTC ACGCTCTGCT CGTTGACACG GTACTGAGTC 8520

GCTACATTCA TCATAACGAA CACATTGTCC TTGGTCTTAG TCTCAACCAC AATATCACTT 8580

TGCAACAAGC GCAACTGAAT CCGTGCTGCA ATCGAGTCAA TCCCAAAAGG CAAGCGAATA 8640

TGAATACCGC TATTAGCAAC CTTTTGGTAT TTCCCAAAGC GTTCAATAAT CGCCACCGAC 8700

TGCTGACGAA CCACATAAAC TGTACTCAGT GTGACTATCA CCAATAGGAG CACACAAACA 8760

ATCAGAAAAA TCATGAAAAA TATTGCCATA ATGGAACCTC CACAAGTATT TTTCTAGTAT 8820

TATAGCACAT TTAAAGAAGG CTGTGCCGTT TTTACTGCGA TTTTTCCTGA AATGTCAATA 8880

ATTAGAGGTG AATTGTCCTA TTGTCGTCCA ATCTCTTGCT AAAATAACTC TTTATAAAAG 8940

GCAATCGTTT CTTCTAAGGT TGGCATAAAT GGATTTCCTG GTGCGCAGGC ATCAATCAAG 9000

GCATTCTTAG AAAGGTATTC AAAGTCGAAA TCTTTTTCTT CAATACCAAG TTCAGTCAGT 9060

TTCTTAGGAA TACCTACTGT CTCAGAAAGC TTCTCAATCT CAGCAATCGC ATAATCGGCA 9120

CATTCTTGAT CTGATTTACC TTCTACATGA AGTCCCAAGG CTTTGGCAAC ATTGCGGAAA 9180

GCTTCTGGTA CACGTTTAGC ATTTTCACGT TCTATAACTG GTAGCAACAT GGCACAGCAC 9240

ACGG 9244

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8898 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

GATCTGAACT TTATCATCAT AACTTAATTT CATAATAAAA ACACCCCAAA AGTTAGATTT 60

TTTCTGTCTA ACTTTTGGGG TGTAGTTCAG TCATTGGACT GACGTTTTTT TGTATGCTTA 120

TTTTGATTTG ATGTAGTTGA TACCATCTGC TTTTGGTGCG ACTGCTTTTC CAAAGAAGGC 180

TGCTAAGACA AGAATTGTCA AAACATAAGG TGCAATTTGA AGATAAACCG CTGGCACTCC 240

TTGTAGGAAC GGCAATTGAG AACCGATAAC AGCCAAACTT TGTGAAAGTC CAAAGAAGAG 300

ACTAGAAAGC ATAGCACCGA TTGGATTCCA TTTCCCAAAG ATCATCGCAG CAAGGGCGAT 360

AAATCCAGGT CCAACAATAG TTGTCACTGA GAAGTTAACT GAGATTGATT GCGCATAAAT 420

CGCTCCGCCA ATTCCACCTA GAAAACCTGA AATAATAACC CCTAAATATC TCATCTTGTA 480

GACGTTGATT CCCAAGGTAT CCGCTGCTTG AGGATGTTCA CCGACAGAGC GGAGACGAAG 540

ACCAAATTGA GTCTTAAAGA GAATAAACCA AGCAAGGAAT GAGAAGGCAA TCGCCAGATA 600

ACCAAGTAGA CTAGTTGACT TGAAGAAGAT ATCACCAATC ACTGGGATAT TTGCCAAGAC 660

TGGGAAATCA AAGCGTCCAA AAGTTTGACT TAGGTTGTCG GTTTGTCCTT TGTTATAAAG 720

AACTTTAACT AAGAAAACAG CCAAGGCAGG CGCCATCAAG TTCAATACCG TACCGCTGAC 780

AACATGGTCT GCACGGAAAT GAACCGTCGC TGCTGCGTGG ATGATAGAGA AAACACTACC 840

AACCAATCCT GCTACAAGCA AGGATAGCCA TGGAGTTGCT GCTCCAAATT GTTCTGCAAA 900

TTCAAGGTTA AAGACAACTC CAGAAAAGGC ACCCATAACC ATAATTCCTT CAAGGCCAAC 960

GTTTACCACA CCACCACGTT CAGAGAAAAC ACCACCGATA CTTGTAAAGA TGAGAGGTGC 1020

TGAGTAAATC AGCATAGAAG ACACCAAGAG GGGGAGCAAG GTTATAATAG ACATCTTTAC 1080

TTACCTCCTT TAACTTGTTT TTTCGGTTTG ACAAAGCGTT CGATAAGGTA ATGAACACTG 1140

ACAAAGAAGA TAATAGACGC TGTTACAATG CTGACAAGCT CAGATGGTAC CTGCGCCGCA 1200

TTCATACCAG GAGCCCCAAC TTGGAGAACG CCAAATAGGA AGGCTGCAAA GAGTATACCA 1260

ATTGGTGAGT TGGCCGCAAG CAAACTAACC GCCATTCCGT TAAATCCGAT AGCTAATGAC 1320

GAACCTTGAA CATAGACGTT CTGGAAGGTT CCCAAACCTT CAACAGCTCC ACCAAGACCT 1380

GCCAAGGCAC CTGAAATAAT CATAGATAGG ATAATAGTCC GCTTGGCAGA AATACCAGCA 1440

TATTCTGAAG CATGTGGATT AAGACCAACT GCACGGATTT CAAAACCAAG AGTTGTTTTC 1500

TTGAGCATGA ACCAAATAAC TGCAACGGCA ATGATGGCAA AGAAAATACC AATATTCATC 1560

CGTGAGTTAC CAGTCAACTC AGCCAACCAA GGTGTCTGAT AGGTTGCATT AGCCCCAACA 1620

CGAATGGTCG AATCTGTACT TTGCATGAAG TCTTTAGGGA AAGCATGGAT AAAGGCATTC 1680

CCTACATACA AGACAATGTA GTTCATCATG ATGGTTACAA TAACCTCTGA CGTCCCTAGA 1740

TAGGCCCTAA GAATACCTGG AATCGCTCCG ACAATCCCAC CAGCAATCAA GGCAATCACG 1800

ATGGTTGCTA GAATCATCAA GGGACGGGGC ATATCTGGAT GCGACAGGGC AAACCAACCA 1860

CTGAGAATCC AACCTGCCAA AGCCTGACCA GGAAGTCCGA CGTTAAAGAA ACCAGCTCGA 1920

CTGGCAACGG CAAAACCAAG ACCAATCAAG ACCAGAGGAC CCATAGCACG GAAGATTTCT 1980

CCAATCCCAC GCAGACTGCC AAAGGCTGTA TAGAACAATT CTTCGTAGCC CCAAATAGCA 2040

TCATAACCGA AGATCCACAT GACAATGGCT CCGAGTAAAA TTCCTAGGAA TACAGAAATC 2100

AAGGGAACCG AAATTTGTTG TAATTTTTTA GACATCACTC TTCTCCTTTC CCAAGTTTCC 2160

ACCAGCCATC AAGACACCAA GTTCTTGTTT ATTGGTTGTT TCTGGTGATA CAATACCTTG 2220

AATCTTACCA TCGTGGATAA CGGCAATACG GTCTGAGACG TTTAAAATCT CATCCAATTC 2280

AAAGCTGACA ACAAGGACAG CCTTGCCATT ATCACGCTCT TCAATCAAGC GTTTGTGGAT 2340

ATACTCAATG GCACCGACAT CCAACCCACG AGTTGGCTGG CTAACGATAA GGAGATCAGG 2400

ATCTCGATCA ATTTCACGAG CAATAATTGC TTTTTGTTGA TTTCCTCCTG AGAGTGCAGC 2460

TGCAGGAACT AATTCACTGG CAGCGCGAAC ATCAAACTCT TCCATCAGCT TTTTAGCATA 2520

AGAAGTAATA TTTGAATAAT TCAAAATTCC ATTTTTACTA TGTGGTTCTT TATAGTAGGT 2580

TTGAAGGGCA ATATTTTCAG ATATCATCAT TTCCAAAATC AAGCCATCAC GGTGACGGTC 2640

TTCTGGAACG TGCCCAACAC TTAGTTCTGT AATCTGACGT GGGTGCAAGC CTACAATTGA 2700

ATCTCCTTTT AGCTCAATGC TACCAGATTC AACCTTACGA AGACCTGTAA TGGCTTGAAT 2760

CAGTTCAGAC TGACCATTTC CATCAATCCC CGCAATACCA ACAATCTCTC CAGCACGAAC 2820

ATCCAAGGAC AGATTTTTAA CAGCTGGAAC ACCACGGTTT TCATTGACCA CCAAATCTTT 2880

GATAGACAAA ACCACTTCTT TTGGTTTAGA GGCTTGCTTC TCTGTTTTAA AGGAAACAGA 2940

ACGTCCTACC ATCATTTCCG CCAAATCAGC ATTGGTAGCC CCTGCAATTT CAACGGTTTC 3000

AATTGATTTC CCACGACGGA TAACTGTAAC ACGGTCAGAA ACTGCTCGAA TTTCATCCAA 3060

TTTGTGGGTA ATCAAGATAA TTGATTTTCC TTCTTTGACA AGATTTTTCA TAATAGCCAT 3120

CAACTCATCA ATTTCTGATG GAGTCAAAAC AGCCGTTGGT TCGTCAAAGA TAAGGATATC 3180

AGCCCCCCGA TAAAGTGTTT TTAAAATTTC TACACGTTGT TGGGCTCCAA CTGAGATATC 3240

TGCTACCTTG GCAGAAGGGT CAACAGCTAA GCCATAACGT TCAGAAAGAG CCTTGATTTC 3300

TTTGCTAGCT CCAGCGATAT CTAGCACACC ATTTTTAGTC AATTCACTAC CTAAAATGAT 3360

GTTTTCAGCC ACTGTGAAGG CTTCAACCAA CATAAAGTGC TGGTGAACCA TCCCGATTCC 3420

CAAGCTAGCT GCTTTAGATG GGGAGTCGAG ATTGACAACT TGACCGTTGA CCGCGATTTC 3480

ACCACTAGTT GGTTCAAGAA GGCCTGCTAA CATGTTCATT AGCGTGGACT TACCAGCCCC 3540

ATTTTCTCCT AAAAGTGCAT GAATTTCACC TTTTCGTAGG TGCAAGTTGA TTTTGTCGTT 3600

GGCAACAAAT CCACCAAACA CCTTGGTAAT ATCACGCATC TCAATGACAT TTTCGTGTGC 3660

CATGTGCTCT TCCTTTCAGA GTCTTATTTT ATTTCAATAA AACTTGCTAG TTTGTCTAGT 3720

AGCAAGCTTT ACTTAGACAA AATGACTTTG TCTCAACTCT TAAAAAAGCG GCCCTTGGCC 3780

GCTTCCTAAG AAATGACTTC CATCCATTAT TTTTCAGGAA CTTTTACGCT TCCATCAAGG 3840

ATTTTAGCTT TTGCATCTTC GACAGCTTTT TTACCTTCTT CTGAAAGGTT TGTTACTGCC 3900

AAGTCAACCC CTTTATCCTT CAATGAGTAA ACGATCACTT GACCGCCAGG GAATTCTCCT 3960

CTTTCTGCCT TGTTAGAAAT ATCTTTTACA GTTGTACCAA CTTGTTTCAA AGTAGATACA 4020

AGAACAAAGT TTGATTCTTT GCCATCTTTA GAAGTGTATT TACCTTCTGC TTCTTGGTCA 4080

CGATCAACAC CGATAACCCA AACTTTTTCA TTTTCAGGAC GGCTTTCGTT GAGAGATTTT 4140

GCCTCTGCAA AGACACCTGC ACCTGTACCA CCAGCTACTT GGTAAACAAT ATCTGCACCG 4200

GCTGCGTATT GTGCGGCTGC AATTGTTTTA CCTTTAGCCG CATCACCAAA TGAACCAGCG 4260

TAGTCAACTT GGACTTTGAT AGATGGGTCT ACTGACGCAA CACCAGCCTT GAATCCTGCT 4320

TCAAAACGAG AGATAACTTC AGATTCGATA CCACCTACAA AACCAACTTG TTTTGTCTTA 4380

GTTGTTTTTG CTGCAGCCAC ACCTGCAAGg TAACCTGACT CATTATCAGC GAAAGTTACG 4440

CTCGCAACAT TCTTTTGGTC TTTAATCACA TCATCAATCA AGACATAGTT CAAGTCAGTG 4500

TGTTCTTTTG CTGCATCTTT AACTGCATTA TTAAGGGCAA AACCAACACC GAAGATTAGG 4560

TTGTAACTTC CAGCCGCTTG TTGCAAGTTG TTAGCGTAGT CAGCTTCACT TGTTGATTGG 4620

AAGTAAGTGA AACCGTTATC TTTTGAAAGA TTGTGTTCTT TACCCCAAGC CTGCAAACCT 4680

TCCCAAGCTG ATTGGTTGAA TGATTTGTCA TCAACACCAC CAGTATCAGT GACGATTGCT 4740

GCTTTTGTCT TCACATCAGA AGATGAAGCT GCGTTACGAG AAGAGCGGTT ACCACATGCA 4800

GCAAGTCCAA CTGCTGCCAC TGCAACTAGG CCAAGACCTA GCCATTGTTT CTTGTTCATT 4860

ACTGAACCTC CTAAATAAGA TGTGCAACGA TGTTGCAAGT ATGGATTGGT TGGCCACAAG 4920

GACCGTGCCA CTCAGAGAGC GACTCAGACT AGTTTAAGTC TGTAAAAGAG TATGGAAGTA 4980

ATTCCCCGAC CGTCATCTCG ACCGTCGATT TATCTTTTGC GACTAAGGTC ACTTTTAGAT 5040

CTTGTTCAAA AAATTCAGCC ATCACTTGGC GACAAGCACC ACATGGCGAG ATCGGTTTTT 5100

CAGTTTGACC ATAGACAATC AATTCTGAAA ATTCTCTTTG GCCTTCAGAT ATAGCCTTAA 5160

AAATAGCTGT TCTCTCACCG CAATTGGTCA AAGGATAGCT AGCATTTTCA ATATTCACTC 5220

CCGTGTAAAC ACTTCCGTCT TTAGCTACTA AAACTGCTCC GATAGGAAAG TGAGAATAGG 5280

GGACATAGGC ATGTTTGCTG GTTTCAATTG CCAGTTCAAT CAACTCAGTA GTCGCCATCT 5340

GCCAATTCTC CTTTTAAAAT AGCTACCCCA GCTGACGTTC CGATACGGGT CGCACCTGCT 5400

TCGACAAAGG CAAGAGCATC TGCATAAGAA CGAGCTCCAC CGGCGGCCTT GACACCCATA 5460

TCAGATCCAA CTGTTTCACG CATTAATGTA ACATCTGCTA TCGTAGCACC ACCAGTTGAA 5520

AAGCCAGTAG ATGTTTTGAC AAAGTCAGCC CCAGCTTTTT GGGCCAATTG GCAAACAACA 5580

ACTTTTTCTT GGTCTGTCAG AAGGCAAGCT TCAATAATGA CTTTCACTAA CTTATCACCA 5640

CTTGCTTCCA CTACTGCGCG AATATCTGAC TCAACCAAGG CTAAATTACC TGATTTGAGA 5700

GCTCCAACAT TGATCACCAT ATCAATCTCA TCTGCACCAT TTTGGATAGC TTCTTTTGTC 5760

TCAAATGCTT TCACGGCTGA AGTTGTTGCT CCCAAAGGGA AACCTACTAC TGTGCAAACC 5820

TTAACATCTG TGCCTTCAAG TCCTTTTTTA GCATGTTCAA CCCAGGTCGG ATTAACGCAA 5880

ACACTGGCAA AGTCATACTC TCTAGCCTCA GACAACAAAC TATCAATTTG TTTTTTCTTT 5940

GCATCTTGTT TTAAAAGCGT ATGATCTATA TATTTATTTA ATTTCATTTC GGTTTTCCCT 6000

CCATTTAGGA GATGATTTCT ACAATTTCAC GGATTTTTTT CACTTCATCA CTTATTTTAA 6060

CACATTTTTG GAAATCTGTA ACTAGTTGAG GTGGAATTTT TTCATTTGTG TATACTTTTG 6120

CAACAATTTC ACCCTTTTGA ACGGAGTCTC CAATCTTCTT TTCAAAAACA ATTCCTGTTT 6180

CATAGTCCAA GGCATCAGAC TTAACTGCAC GACCAGCACC CAGCCTCATG GCATAAAGAC 6240

CAAAGTCCAT AGCTGGAAGA GCTGAAATGA CACCCGTTTC CTGAGCAGGG ATTTCCACCA 6300

CATGAGCTAC ATTTACAGGA CGATAGAGGT CTTCCAAGTC TCCACCTTGG GCTTGCACCA 6360

TTTCCTCAAA CTTAGCCAGT GCTTGACCAT TCTCAAGATG TTGGTGAACT TCTTCAACAG 6420

TTTTGTTAAC ATTTGCCAAA CCAAGCATAA TTTGAGCCAA TTCACAAATA AAGTGGGTAA 6480

TATCCTGACG TCCTTGACCT TGCAAAATCT CCAATGCTTC AAGGATTTCC AGACGATTTC 6540

CAATCGCTCG TCCCAAAGGC TGGCTCATAT CCGTAATCAC TGCTACTGTC TTCCGTCCAA 6600

CAACCTTACC AAGATCTACC ATAGTTTGAG CCAACTCACG CGCCTCATCA ACCGTCTTCA 6660

TGAAGGCACC CTCACCGACA GTCACGTCTA GCAAAATAGC ATCCGCCCCT GCCGCAATTT 6720

TCTTGCTCAT CACCGAACTC GCAATCAAAG GAATCGTGTC GACAGTTGCG GTCACATCAC 6780

GAAGGGCATA GAGAAGCTTA TCTGCTTTGA CCAGCTGGTC TGATTGCCCA ATGACAGATA 6840

CTCCAATATC CTGAACCTGA CGAATAAAAT CCTCTTGACT ACGTTCTACT TGATAGCCCT 6900

TAATGGACTC CAATTTATCA ATTGTTCCGC CTGTATGGCC AAGACCACGA CCACTCATTT 6960

TTGCTACAGG CACACCGAAG CTAGCAACAA GAGGAGCTAA AATCAAGGTT ACCTTATCGC 7020

CGACACCACC AGTAGAATGC TTGTCAACTT TCACACCATC AATGGCTGAC AGGTCAAACT 7080

CTTGCCCAGT CTTAACCATA TTCATCGTTA AATCAGAGAT TTCTCGAGTC GTCATTCCTT 7140

TAAAATAAAC AGCCATAGCA AAGGCAGACA TCTGATAATC AGGAACAGTT CCTGATACAT 7200

AGCCTTCTAT CAGCCATTCA ATTTCACTTG AAGTCAGTTC TTGACCGTCT CGTTTTTTTT 7260

GGATTAAATC AACTGCTCTC ATTCTTTCAC ACTTCTAAGG ATATAGTATC CCTTGTCTTT 7320

TTTAAGGATT TCACAATTGC CAAACACATC TTCCATCTTA GACTTGGCAC TTGGAGCTCC 7380

TTGTTTTTTC TGGATGACGA TGGTCAAATC TCCACCAATT TCCAAGAAAT CTTTACTTTT 7440

CTCGATGATT TCATGAACGA CTTGCTTGCC CGCACGGATA GGAGGATTGG AAATGACATG 7500

GTCAAATCGC CCTTGAACTC TTGCATAAAT ATTAGATTGA AATATCGTCG CTTTTGCATT 7560

ATTTTTTTCA GCATTTCTCT GAGCTAAATC CAGGGCACGA GTGTTAATAT CAACCATGGT 7620

CGCCTGAACT CCGTAAACCT TGACCAAGGA CAAACCTAAT GGACCATAAC CACAGCCTAC 7680

ATCTAGGACT GTCTCTCCTT GGTTGACATC CAGACACTTG AGCAAGAGTT GACTTCCAAA 7740

GTCAACCATT TTCTTGCTAA AAACACCCGC ATCTGTCAAA AAAGTCATTT TTTCTCCCAA 7800

CAAGTCCACT CTCAACTCAT GAATGTCGTG AGCAGCGTCA GGATTTTCTG CATAGTACAT 7860

TTTACTCATG ACACTATTTT ACCATAATTT GACTCAAATT GTAAATCGTT TACAAATTGA 7920

TAATAAAACG AAAAAGACCG AAGAAAGCAA GTCACGAAGC CATTTTCTTC AATCTCTTTC 7980

AACACTTATA AATAATAAAC CATTTAGAAC TATAAATATC ACAGTCCAGA TAAAAACAAA 8040

AAGTTTATCA TCTATAATCA GGCAGATTAT TATTTCTATT GCTTAACCTT AAAATACTTT 8100

ATTATCAACA AAATTCCTAA CAAAATGTTT AGATAAAAGC CCAACTGATA CGTTTATGTC 8160

AGGATTTCCA AACTTGTCCA AAGTCGTATC AAATCTTCTA GTGACATGTG GAAGAAATAA 8220

CCCTCTGTCG CAATCCGTAG GACTAAAAAG CAATAACTAC CCGCAGCAAT CCATTTCGTC 8280

CATCGTTTTT TAGTAAGAAA GCAATTAAGA ACGAACAAAT AAAGACAGCT GTTACAATAG 8340

CATGTTCCAT CAAAAAAGTA AAACCGTAAT AGGTTTCCAC AAAGCATCTA CCATTATCTG 8400

CATTGGTTCC TTTTATAAAA GGTAAAGCAA AACTTAAAAT AAAACAGAGT TCCAATATGT 8460

AACGTTTTAA GATTTTCATA GTACACCTCC TATAAGTTGT GAACTAAAAA GCCCCCTTTA 8520

TAAGCTTATA AATCAGTAGA ATCTATCTCC TATTTCATCA ATAAATTGAT CACTTATACT 8580

ATATACCATT GACTTACCAC ATTCAAGAAA CCGCTTTATT TTTTTAGCTT TTTATGGTAT 8640

GATAGACAAA ATATCTAGGG GAAAACAAAT GACCAACGAA TTTTTACATT TTGAAAAAAT 8700

CAGCCGCCAG ACTTGGCAAT CTTTACATCG AAAGACAACA CCTCCTTTGA CAGAAGAAGA 8760

ATTGGAATCT ATCAAGAGTT TTAATGACCA AATCAGTCTC CAAGACGTTA CAGATATCTA 8820

TCTCCCCTTG GCTCATTTGA TTCAGATTTA CAAGCGAACT AAGGAAGATT TAGCCTTTTC 8880

AAAAGGAATT TTCCTCCA 8898

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13188 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

TATCTTAACG aGGATTGGGT TTATCGTCAG TCTTATTGCC CTAATTGTGG GAACAATCCC 60

TTAAATCATT TTGAAAATAA TCGGCCTGTA GCAGATTTTT ACTGTAATCA TTGTAGTGAG 120

GAGTTTGAAC TAAAGAGCAA AAAAGGAAAT TTTTCATCAA CAATCAATGA TGGTGCTTAT 180

GCAACGATGA TGAAGCGTGT GCAGGCAGAT AATAATCCTA ATTTCTTTTT TTTAACTTAC 240

ACAAAAAATT TTGAGGTAAA TAACTTTCTT GTCCTTCCGA AGCAATTTGT TACACCGAAA 300

TCGATTATTC AAAGAAAACC ACTTGCACCA ACTGCTAGAC GAGCAGGTTG GATTGGTTGT 360

AACATTGATT TATCACAAGT ACCTTCTAAA GGAAGGATAT TTCTTGTGCA AGATGGACAA 420

GTTAGAGATC CAGAAAAAGT TACAAAAGAA TTTAAGCAAG GTTTATTTTT AAGGAAGAGC 480

TCTCTGTCAT CAAGAGGTTG GACAATAGAA ATTCTAAATT GTATAGATAA GATAGAGGGT 540

TCAGAATTTA CCCTTGAAGA TATGTATCGT TTTGAAAGTG ACCTAAAAAA TATCTTTGTT 600

AAGAACAATC ATATCAAAGA AAAGATTAGG CAACAGCTTC AAATATTAAG AGACAAAGAA 660

ATAATAGAAT TTAAAGGTAG AGGAAAGTAT CGGAAATTAT GAAAACGAAA CAACTTGTTG 720

CATCAGAAGA GGTGTATGAT TTCTTAAAAG TCATCTGGCC TGATTATGAA ACTGAAAGCC 780

GTTACGATAA CCTAAGTTTA ATCGTCTGTA CCTTATCAGA TCCCGATTGT GTGAGATGGT 840

TATCTGAAAA TATGAAATTT GGTGACGAAA AACAACTAGC TTTGATGAAG GAAAAATATG 900

GGTGGGAAGT AGGAGATAAA TTGCCAGAGT GGCTACATAG CTCCTATCAT AGATTATTGT 960

TAATAGGTGA ATTATTGGAA AGCAATCTAA AACTGAAAAA GTATACAGTA GAAATTACAG 1020

AAACTTTATC ACGTTTAGTA AGTATAGAGG CTGAAAATCC AGATGAAGCC GAACGACTTG 1080

TAAGAGAAAA GTATAAGAGT TGTGAAATTG TTCTTGATGC AGATGATTTT CAGGACTATG 1140

ACACTAGCAT ATATGAATAG GTAGATGTTT TTATTTTGTC AACAAAAAAG AGGCTCGCAC 1200

CTCTTTTTCT TATTTCTTTT TATGATTTAA TACGGCATTG AGGACAATAG CGAGTAGGCT 1260

GGCTACGACG ATTCCGTTTG AGAAGAACAT TTGGAAGGCT GTCGGCATGC TGACAAAGAG 1320

ATTACTGTTG TTGAGACCGA CACCTGCAGC GATTGAAACA GCTGCGATAA GGAAGTTGTG 1380

TTCATTGTTA GCAAAGTCAA CACGGGCGAG GATTTGCATC CCTTGAATTG ATACAAAACC 1440

AAACATTACC AGCATGGCAC CACCGAGGAC GGAGCTTGGA ATGATTTGGG CAAGGGCGCC 1500

AAACTTAGGA AGCAGTCCAA GGAGAACCAG GAAACCAGCT GCGTAGTAGA TTGGCAGGCG 1560

TTTTTTGATG CCTGACAATT TAACCAAACC AACGTTTTGT GAAAATCCGG TGTAAGGGAA 1620

GGTGTTAAAG ATTCCTCCGA GAAGTACGGC CAAACCTTCT GCGCGGTATC CGTTGCGAAG 1680

GCGCGTGCTG TCGATTGGAT CCTTTGTGAT ATCAGACAAG GCCAGATAAA CACCAGTTGA 1740

CTCAACCATA GACACCGTTG CGATGATACA CATCATGACA ATAGATGAGA TTTCAAAGGT 1800

TGGCATCCCA AAGTAGAGTG GAGTTGGGAC ATGGACAAGT GGAGCTACCG CAACAGGAGA 1860

GAAGTCCACC AAGCCCATAG TAGCAGCAAT GGCAGTTCCA ACAACCAGAC CAATCAAAAT 1920

AGAGATAGAC TTGATAAATC CTTTGGTAAA GATGTTGATC AAGAGGATAA TCAGAACAGT 1980

AATAGCTGCA AGCAAGAGAC TTTGACCAGT TGGCTCTGGA ACGTTATTTC CCATATTTCC 2040

AATAGCGACA GGGATCAAGG TTAAACCAAT CGTGGTAATA ACAGATCCTG TTACGATAGA 2100

TGGGAAGAGA TTGGCTACTT TTGAGAAGAT GCCTGAAACA AGAACCACGT AAATCCCAGA 2160

TGCGATAAGG GCACCAAACA TAGCGCCACT ACCATGGCTT TGCCCAATCA TAATCAAGGG 2220

AGCGACCGAC TGGAATGCAA CTCCAAGAAC GACTGGGAGT CCAATCCCAA AGTATTTGTT 2280

GAGTTGGAGT TGGAGGAAGG TTGCCACCCC ACACATGAAG ATATCTGTAG AAATCAGGTA 2340

GGTCAACTGC TCAGCTGAAT AGCCAAGGGC TGTCGCAATC ATGATGGGAA CCAGGATAGA 2400

TCCTGAGTAC ATGGCTAGTA AGTGCTGCAA GCCAAGAACG GCTGCTTGCG AGTGTTTTTC 2460

TTGAGTTTGC ATTAGAGATC TGCCTCCTTA AATACGACTT GACCATTTTC AAAACAATCC 2520

AAACGAGCAA GTGATAGGAC AGGGTAGCCT GCTTTTTCAA GCAAATCACG ACCATCTTGG 2580

AAGGATTTCT CAATCACGAT ACCGATAGCT TGGACTGTGG CACCGGCCTG TTCGATGATT 2640

TGAATCAAGC CTTTAGCAGC TTGGCCATTA GCAAGGAAAT CGTCGATAAT CAAAACCTTG 2700

TCCTCTGGTG AGAGGAATTT TTCAGCGATA GAAACGGTGC TGGTCACCTG CTTGGTAAAG 2760

GAGTAGACTT GAGCAGTTAA GATGCCTTCG TTCATGGTGA TGTTCTTAGC TTTTTTGGCG 2820

AAAATCATGG GAACGTTTAA GGCTTCAGCT GTAAAAACGG CTGGGGCAAT ACCCGACGCT 2880

TCAATGGTTA CGACCTTGGT AATGCCAGTA GTAGCAAATT TTTCCGCAAA AACCTTACCA 2940

ATCTCTCGCA TCAAGCTAAA GTCAACTTGG TGGGTTAAAA AGGAATCTAC CTTGAGGATG 3000

TTATCACCCA AGATATGCCC ATCCTTGAGG ATGCGCTCTT CTAATAATTT CATAAGACCT 3060

CCTAAAGTCT AAAAGTTAAT TTACTTGTTG TTTAAATATT TCTATAGTGA TCCCTTTTGC 3120

TAATACTATA TATTTGATAA AACTATTACG AGCGAAGCGA GTCTTATCAA ATATTTCCCG 3180

TTGTAGTGGT ATCATAGACA ATAATCTTGT TATTGTCTAT GACGGGATTT TTGAGAGTAA 3240

AATAGTTCGG GGAACTATTT TAGCCTAAGC CTAGAAATGA AAGAGCTAGG GGCTCAAAAA 3300

TTAGGGATGA AATTCCCTGG ATTCCTGAAA TTATTCACAG GATAATTTCA CCTCCCGTCC 3360

GCACTAATTA AGGGAAATAT TAAAAAAAGA CCTACTTAAT CTCTAAGTAA GTCCCCTAAA 3420

TAGACATGGC AAAAACGGCC ATATCTCACT GCTGACTTAC TTATTGTTAG GTGTTCCGGC 3480

ACCTTGTAGA AACGTCGTGC CAATTCACGA CATAAACAAG TAAAACGATA TTCAATTTTA 3540

AATAGGCTTG AGCCAATGTT TTTATTTTAC ACTAAATAAC TTTAGAAATC AACTATTTTG 3600

TTAGTGTTTT GGTTTAAAAA ACGAACAAAA AGAAGAGAGG GTGAACAAAA ACTCCATTGT 3660

AAGCTAACAG TTATACTAAA TGAAAATCAA AGAGCAAACT AGGAAGCTAT CCACAACCTC 3720

AAAACACTGT TTTGAGGTTG TGGATAGAAT TGACAGAGCC AGTATCATAT ACCTACGGTA 3780

AGGCGACGTT GACGTGGCTT GAAGAGATTT TCGAAGAGTA TTAGAAGATT TTTCCATCAT 3840

AAAAGGCATA CTATCAAGCT TTTAGACACC TGACAATATG CCTTTTTCTA ACTTTAAAGA 3900

CTTTTCCCAA TTTTTATTAT TCTACTCGCT AAATCTTAAA AAATAGCCAT CTGGATCCAA 3960

AACTGCAAAT TTATGAGGAT AGATATAGGG ATCACTGACA CGAAACTTTC TTTTGGTCAA 4020

GGGACGATAA ATAGGATAGT TTGCCTTCAT CACTCTTTAA TAGAGTTTTG AAACATCCTT 4080

TATGCCAAAG GAGAGATTGA CTCCACGACC AAAGGGATAG GTCAGTTCAG CTAGTTGATC 4140

CTTTGTTCCC TCCTCTAACA TTAGTTGACA CTCTTCAAGA GAAAGAGAAA GTTTTCTTCT 4200

GGACGTTGGT ATTCAATCCT AAAACCCAGT AAACCACAGT AGAAGGACCG GGACTGTTCG 4260

ATATTCGATA CAAGCAACTC GGGAATGACC GCATTGTAGT CCATATAGAA AATCCTTACA 4320

AGTCAATTTC CAAGACAATC GGTGTATGGT CTTGGCGAGC ACCTGAGTCA ATCATATCAG 4380

ATTTAGTGAC CTTGTCAGCG ATACGGTTAC TTGTGAGCCA GTAGTCGATT CTCCAGCCTG 4440

TATTGTTGAT TTTAGAAGTT TTGCTGCGTT GTGCCCACCA AGTGTAGCGT TCAGGAACAT 4500

CGCCATGAAC ATGGCGGAAG GTGTCTGTAA ATCCAGTTGC CAAAAGGTTG GTAAATCCAG 4560

CACGTTCCTC GTCAGTAAAT CCAGGTGAAC GGCGGTTGCT AGCAGGATTT GCAAGGTCGA 4620

TTTCATTGTG GGCTACGTTG TAGTCACCGG TCGCAAGGAC TGGTTTTTCT TTGTCTAGTT 4680

CAGCCAAATA CTCAGCATAT TTGGCATCCC AGACTTGGCG TTCTTCCAAG CGTTTGAGAC 4740

CGTCACCAGC GTTTGGAGTG TAAACTTGGG TTACGAAAAA TGCATCAAAT TCTAGAGTGA 4800

TGATACGACC TTCCAAGTCC ATGGTAGAAG GGGCACCGAT TTCTGGGAAG CTGATAGTAG 4860

GTGTAAGTTC TTTCTTATAA AGGAACATGG TTCCAGCATA GCCTTTACGG GCAGGCTCTT 4920

GGGAAGAGCG CCACGTGTTT TCGTAGCCTG GGAAGAGTTC TTCTAAAATT TCCACGTGTT 4980

TCTTTGTAGG TCCTTTGGCA GAAAGCTTGG TTTCTTGGAT AGCAATGATA TCAGCATTTT 5040

CAGCGACCAA GGTTTGTAGG ACTTCTTGGG ACAATTTGGC ACGAGCTGAG TCACTAGTTA 5100

GGGCAGCGTT TAGGGAATCA ATATTCCATG AGATAAGTTT CATAAAGTTA CCTTTTTCAT 5160

TCAGATTATA GATTTTATTA TACCAAAAAA AGATCTATTT CCCCAACGTA TGGTTTGAAA 5220

AATTACTCTC TTTCGTTTAT AATTAAGAAT GATTTTATGA AAGGGAGTGA AAATACATGA 5280

AATTCTACTC TTATGACTAT GTACTCAGCC AAATCGGTCA GCAAAATGGT ATCATGGTTG 5340

GCTTTGGGAT TGTTCTATTA GCTGTGACAG TTTTTTTTGC TTTCAAGGCA TACCATAATA 5400

AAAAGGGAAG CGAATTTCGT GAGTTGGTCA TGATTTCAGA TCTGGCCTTA TTTAGCTCTG 5460

CTTTTGGTCA GCATCACGAC TTATCAAAAC AATCAAGTTT CTAACAATAA ATTTCAAACT 5520

TCACTTCATT TCATCGAGGT TGTTTCCAAA GATTTGTGAG TAGACAAGTC AGAAGTCTAT 5580

GTTAATACTT CCACAAACAC AGATGGCGCA CTTATCAAGG TGGGAGATCG CTATTATCGT 5640

GCCCTAAATG GAAGTGAGCC AGACAAGTAC CTGTTAGAGA AAGTCGAATT GTATAAGACA 5700

GACGCAATTG AACTGGTGGA TGTGAACAAA TGACACTTAA TTATATCGAA ATTTTAATCA 5760

AACTGGTCTT GACTCTCAAA TAGCTCAACA ACAATGTTCA CTTTGTGAAA CGTTTGATTG 5820

ATGGTAAGCC AACTCTCCTT ATCAAAAATG GGAATATTGA CCCAGAAGCC TGTCGTTCAG 5880

TTGGTTTGTC TGCATCGGAT GTATCCCTCA AACTTCGTAG CCAAGGGATT TTCCAGATGA 5940

AGCAAGTCAA ACGAGCTGTG CAAGAGCAAA ATGGGCAACT CATCGTTGTG CAAATGGGAG 6000

ATGAAAATCC TAAGTATCCA GTTGTGACTG ACGGTGTGAT TCAAGTAGAT GTCTTGGAAT 6060

CGATTGGTCG TAGCGAAGAG TGGTTGCTTG ATAACCTCAG TAAACAAGGG CATGACAATG 6120

TAGCCAATAT CTTTATTGCT GAATATGACA AGGGTGCTGT TACAGTCGTA ACTTATGAAT 6180

AAGAAAAACC TGGGGTCTTG TACTCTTCGA AAATCTCTTC AAACCGCGTC AACGTCGCCT 6240

TGCCGTATGT AGGTTACTGA CTTCGTCAGT TCTATCTACA ACCTCAAAGC AGTGCTTTGA 6300

GCAGCCTGCG GCTAGTTTCC TAGTTTGCTC TTTGATTTTC ATTGAGTATT GGCCTCAGGT 6360

TTCCATTTGC AATCAGAAAG GGATTTTATG TCCATTATTC AAAAACTTTG GTGGTTTTTC 6420

AAGTTAGAAA AACGCCGTTA TCTAGTCGGA ATTGTGGCCC TGATCTTGGT TTCCGTCCTC 6480

AATCTCATTC CTCCTATGGT TATGGGGCGG GTCATTGATG CCATCACATC GGGGCAATTA 6540

ACCCAGCAGG ACCTCCTTCT TAGCCTATTT TACTTGCTAC TTGCAGCCTT TGGTATGTAC 6600

TATTTGCGCT ATGTGTGGCG TATGTATATC CTTGGGACCT CTTATTGCTT GGGACAGATC 6660

ATGCGGTCTC GCTTGTTTAA GCATTTCACA AAAATGTCGT CAGCCTTTTA TCAAACCTAT 6720

CGGACGGGTG ATCTGATGGC ACACGCAACC AATGATATCA ATGCCTTGAC TCGTTTAGCA 6780

GGTGGCGGTG TCATGTCTGC GGTGGATGCC TCTATCACGG CTCTGGTGAC TTTGTTGACC 6840

ATGCTCTTTA GCATCTCATG GCAGATGACT CTTGTTGCCA TTCTCCCCCT ACCTTTCATG 6900

GCCTATACGA CTAGTCGCCT AGGGAGAAAG ACTCATAAGG CCTTTGGCGA ATCCCAAGCT 6960

GCTTTTTCTG AACTCAATAA CAAGGTACAG GAGTCCGTAT CAGGTATCAA AGTGACCAAG 7020

TCTTTCGGTT ATCAGGCAGA CGAGTTGAAG TCTTTTCAGG CAGTCAATGA ATTAACCTTC 7080

CAAAAGAACC TGCAAACCAT GAAATATGAT AGTCTCTTTG ACCCTATGGT TCTCTTGTTT 7140

GTTGGTTCGT CCTATGTTTT AACGCTTTTG GTTGGCTCCT TGATGGTTCA GGAAGGGCAG 7200

ATTACAGTTG GGAATCTAGT CACCTTTATC AGCTATTTGG ATATGCTGGT CTGGCCTCTT 7260

CTGGCCATCG GTTTCCTCTT TAATACTACT CAGCGAGGGA AGGTTTCTTA CCAGCGGATT 7320

GAAAATCTTT TGTCTCAGGA ATCTCCTGTA CAAGACCCTG AGTTTCCTCT GGATGGTATT 7380

GAAAATGGGC GTTTGGAGTA TGCCATTGAC AGCTTTGCTT TTGAAAATGA GGAAACACTG 7440

ACGGATATTC ACTTTAGTTT GGCAAAAGGG CAAACACTGG GCTTGGTTGG GCAGACAGGC 7500

TCTGGGAAAA CGTCCTTAAT CAAGCTCCTC TTGCGTGAAT ACGATGTGGA TAAGGGTGCC 7560

ATTTATCTAA ACGGTCACGA TATTTGGGAC TATCGTCTGA CAGACCTTCG CAGTCTCATG 7620

GGCTATGTTC CTCAGGACCA GTTTCTTTTT GCGACTTCAA TCCTAGACAA TATCCGCTTT 7680

GGCAATCCTA ACTTGCCCCT TTCAGCGGTC GAGGAAGCTA CTAAGCTAGC CCGGGTTTAC 7740

CAAGATATTG TAGACATGCC TCAAGGATTT GATACGCTGA TTGGTGAAAA AGGAGTCACT 7800

CTTTCTGGTG GTCAAAAGCA ACGGTTGGCT ATGAGTCGGG CTATGATTTT AGACCCTGAT 7860

ATCTTGATTT TGGATGATTC CTTATCCGCC GTAGATGCCA AGACAGAGTA TGCGATTATC 7920

GACAACCTCA AGGAGATGCG AAAGGACAAG ACAACCATTA TCACTGCCCA TCGCCTCAGT 7980

GCTGTTGTCC ATGCAGATTT TATTTTAGTT CTACAAAATG GTCAAATTAT CGAACGAGGC 8040

ACGCACGAAG ACTTGCTAGC TTTGGATGGC TGGTATGCCC AAACCTACCA GTCTCAGCAG 8100

TTGGAAATGA AAGGAGAAGA AGATGCAGAA TAAACAAGAA CAATGGACTG TATTGAAGCG 8160

CTTGATGTCT TATCTCAAGC CTTATGGACT CCTGACCTTT TTGGCACTCA GTTTTCTCCT 8220

AGCGACGACG GTCATTAAAA GTGTCATACC CCTCGTGGCT TCCCACTTTA TCGACCAGTA 8280

TCTCAGCAAT CTTAACCAAC TAGCCGTTAC CGTTTTGCTG GTCTACTATG GTCTCTACAT 8340

CCTACAAACT GTAGTTCAGT ATGTCGGCAA TCTTCTCTTT GCGCGCGTGT CTTACAGTAT 8400

TGTTAGGGAT ATTCGTCGGG ATGCCTTTGC CAATATGGAG AAACTGGGCA TGTCTTACTT 8460

TGACAAGACG CCAGCAGGTT CTATCGTTTC TCGTTTGACC AACGATACCG AGACGATTAG 8520

TGATATGTTT TCTGGGATTT TATCCAGCTT TATCTCAGCA GTTTTTATCT TTCTGACAAC 8580

CCTTTATACC ATGTTGGTGC TGGATTTTCG TTTGACGGCT TTAGTCTTGC TCTTTCTTCC 8640

TTTGATTTTC CTTTTGGTCA ATCTCTATCG AAAAAAGTCA GTGAAAATCA TCGAGAAAAC 8700

CAGAAGTCTC TTGTCAGATA TCAATAGTAA GCTGGCAGAG AATATCGAGG GAATCAGGAT 8760

TATTCAGGCC TTTAATCAAG AGAAGCGCCT GCAGGCAGAA TTTGATGAAA TCAACCAAGA 8820

ACACTTGGTC TACGCCAACC GTTCTGTAGC CTTGGATGCC CTCTTTTTGA GACCTGCCAT 8880

GAGTTTGCTG AAACTTCTAG GCTATGCAGT CTTGATGGCC TACTTTGGCT ACCGTGGTTT 8940

TTCTATCGGG ATAACGGTCG GGACCATGTA TGCCTTTATC CAGTACATCA ACCGCCTTTT 9000

TGACCCCTTG ATTGAGGTGA CGCAAAACTT TTCAACTCTG CAAACGGCTA TGGTTTCTGC 9060

AGGTCGTGTC TTTGCCCTGA TAGACGAGAG GACCTATGAA CCTCTTCAAG AAAATGGGCA 9120

AGCCAAAGTC CAAGAAGGCA ATATCCGTTT TGAACATGTG TGTTTCTCAT ATGACGGTAA 9180

ACATCCGATT CTGGATGACA TTTCTTTCTC TGTTAATAAG GGTGAAACCA TTGCCTTTGT 9240

AGGTCATACA GGTTCAGGGA AATCGTCTAT TATCAATGTC CTCATGCGCT TTTATGAATT 9300

CCAGTCAGGG AGAGTTCTCT TGGATGATGT GGATATCAGG GATTTCAGTC AAGAAGAGCT 9360

GAGAAAAAAC ATCGGTTTGG TCTTGCAGGA ACCCTTCCTC TATCATGGAA CTATTAAGTC 9420

CAATATCGCC ATGTACCAAG AAACCAGTGA TGAGCAGGTT CAGGCTGCGG CAGCCTTTGT 9480

GGATGCAGAT TCCTTTATTC AAGAACTTCC TCAGGGGTAC GACTCCCCTG TTTCCGAGCG 9540

TGGTTCGAGC TTCTCTACTG GGCAACGCCA GCTTCTTGCC TTTGCTAGAA CAGTCGCCAG 9600

CCAGCCTAAA ATCCTGATTT TGGATGAAGC GACAGCCAAT ATTGACTCTG AAACAGAAAG 9660

CTTGGTTCAA GCTTCTCTGG CGAAGATGAG ACAGGGCCGA ACAACTATTG CTATCGCTCA 9720

CCGCCTTTCT ACTATTCAAG ATGCCAACTG CATCTATGTC TTGGATAAGG GACGCATTAT 9780

CGAGAGTGGA ACCCATGAGG AACTCTTGGC TCTGGGAGGA ACCTATCACA AGATGTATAG 9840

TTTGCAGGCA GGGGCCATGG CCGATACTCT TTGAAAATCT CTTTAAACCA TGTCAGCTTT 9900

ATCTGCAATC TCAAAGCTGT ACTTTGATTT TCATTGAGTA CTAGAAGGAA ATCCTTCAAA 9960

TTACAGATTT CTTTCACCGC CTTTTCCATT TTGTGGTATA ATGAAAAATG TTGACAAATA 10020

GTATAATAAA AACAAAGGAG AACAGCATGC TGAAATGGGA AGACTTGCCT GTGGAAATGA 10080

AATCAAGCGA GGTTGAGTCT TACTACCAGC TTGTCTCTAA AAGGAAGGGT TCGCTGATTT 10140

TCAAGCGTTG CTTGGACTGG GTTTTGGCCT TGGTCTTACT GGTTCTGACC TCTCCCATCT 10200

TTCTCATCTT GAGCATTTGG ATCAAGTTGG ATAGCAAAGG GCCAGTGATT TACAAGCAAG 10260

AGCGTGTGAC CCAGTACAAC CGTCGGTTCA AGATTTGGAA GTTTCGTACC ATGGTGACGG 10320

ATGCGGATAA AAAAGGAAGT CTGGTGACTT CTGCTAACGA TAGCCGCATT ACCAAGGTTG 10380

GAAATTTCAT CCGACGTGTC CGTTTGGACG AACTGCCTCA GTTGGTCAAT GTCCTTAAAG 10440

GTGAGATGTC CTTTGTCGGT ACACGACCTG AAGTGCCACG TTATACAGAG CAGTATAGCC 10500

CTGAAATGAT GGCAACCTTG CTCTTGCAAG CAGGGATTAC CTCTCCAGCC AGCATCAACT 10560

ACAAGGATGA GGACACAATT ATCAGTCAAA TGACGGAGAA AGGTCTGTCA GTTGATCAGG 10620

CCTATGTGGA GCATGTTCTT CCTGAAAAGA TGCGCTATAA CCTCGCCTAT CTCCGAGAGT 10680

TTAGTTTCTT TGGGGACATC AAAATCATGT TTCAAACCGT GTTTGAGGTA CTAAAATAAA 10740

GTAGTCATAA GAAAATGAGT ACAGATAAAA GGAGCAAATC AATGCCAAAT TACAATATTC 10800

CATTTTCACC GCCTGATATC ACAGAAGCAG AAATTACTGA AGTAGTGGAT ACCCTGCGTT 10860

CTGGTTGGAT CACAACAGGT CCTAAAACAA AAGAACTGGA GCGCCGCTTG TCTCTTTACA 10920

CACAGACACC TAAGACTGTT TGTCTCAACT CTGCGACAGC CGCTCTGGAG TTGATTTTAC 10980

GCGTTTTGGA AGTGGGACCT GGTGATGAAG TCATCGTTCC AGCCATGACC TATACGGCTT 11040

CATGTAGTGT CATTACGCAC GTGGGAGCAA CCCCTGTCAT GGTGGATATC CAAGCAGATA 11100

CGTTTGAGAT GGACTATGAC CTGCTTGAGC AAGCTATCAC TGAGAAAACT AAGGTGATTA 11160

TTCCAGTAGA GCTCGCAGGG ATTGTTTGCG ATTATGACCG TTTGTTCCAA GTCGTGGAGA 11220

AAAAACGTGA CTTCTTTACC GCTTCAAGCA AGTGGCAAAA GGCCTTTAAC CGTATTGTCA 11280

TTGTCTCTGA TAGTGCCCAC GCTTTGGGAT CTATTTATAA AGGACAACCT TCTGGTTCTA 11340

TCGCTGACTT TACTTCCTTC TCATTCCATG CAGTTAAGAA CTTTACAACG GCAGAAGGTG 11400

GAAGTGCGAC TTGGAAAGCC AATCCAGTGA TTGATGACGA AGAGATGTAC AAGGAATTCC 11460

AAATCCTTTC CCTTCACGGG CAAACTAAGG ATGCTCTTGC CAAGATGCAA CTGGGGTCAT 11520

GGGAATACGA TATCGTTACA CCAGCCTATA AGTGCAACAT GACCGATATC ATGGCTTCAC 11580

TTGGTTTGGT ACAATTGGAC CGCTATCCAA GTTTGTTGCA ACGCCGTAAG GACATTGTGG 11640

ACCGCTATGA TAGTGGTTTT GCAGGTTCTC GCATCCATCC TTTGGCACAC AAGACTGAAA 11700

CTGTCGAATC TTCACGCCAC CTCTACATCA CCCGTGTAGA AGGAGCAAGC CTAGAAGAAC 11760

GCAACCTCAT CATCCAAGAA TTGGCTAAAG CAGGAATTGC AAGTAATGTT CACTACAAAC 11820

CGCTTCCTCT CTTGACAGCC TATAAGAATC TTGGATTTGA TATGACGAAC TATCCTAAGG 11880

CCTATGCCTT CTTTGAGAAT GAAATTACCC TCCCTCTTCA TACTAAATTA AGCGATGAAG 11940

AAGTAGACTA TATCATTGAG ACTTTCAAAA CAGTTTCTGA AAAAGTGCTA ACTTTATCAA 12000

AAAAATGACA AACTACAGTC AAGCGAAAGT GATCCTGCCC CTAAAAAGTC TAATTGAGTG 12060

TAAAAACTGT TGTTTTCAAT TGATAATAGT TTACACCTGT AGTTGAGGCC CCTTTCTCCT 12120

CAGAGAGAGA ATTTTTATAG GATTTTCCTT TCTTGTGGGA GTCCCGTGGT TTGAAATAAG 12180

ATGTGAGCAA TTTAGTGTAG CATTTAGAAT CCTTACTAGA CATCATTTAG AAAATCTAGT 12240

GTCTTGTTCT AGTTTTCAAT TCACCCTATT TTTTGAAAGA CGTGAGTTTC CATGAGTGAG 12300

ATTGTGGAAA CTCGCGTCTT TTTTTGTTTT CAGAATATTG TTCAAAATTT TGTGCCTGTC 12360

TTTCATGTTC TAGTCATTCT TTTGCATGAT AGAATTTATA GCATGTTGAT ATTATAATAA 12420

TACAAATATT CTATATGTTT AGTGATGCTT GCTATACATT ATTAGATCTC CTGCGAGACA 12480

ATCTATAAAA CACTTGTCTA CGATTACCTA TATGCCCTAT TCCAGTATTT TAGAAGCACT 12540

GCATCTATTT TTATCGAGGT TAAATCTAGC TTTTATAGAA GGTCTATTTA AGAAATATAT 12600

TGTAGTGTTT TAGTTTCAAT CCGCCATATG AGCGATATTC AGGTAAATAT CCCTGGCGAA 12660

TGCTTGTATG ACAAGGTATT TGTTCTTTCA TTTATAATTT ACAACATATC AACAAATTTA 12720

AATATAGTAA ATGGGATATT TTATATTCAA GCTAAGAAAG ATAGCATCAC TTTTGAATGG 12780

AAGGCTAAAG AGCAAACTAG GAAGTTGGCC ATAGATAGCT CAAAACCCTG CTTTGAGGTT 12840

GTAGATATAG TAAAATGAAA TGAGAATAGG ACAAATTGAT CGGGACAGTC AAATCGATTT 12900

CTAACAATGT TTTAGAAGTA GAGGTGTACT ATTTTAGTTT CAGTCTACTA TAGAACTGAC 12960

CAAGTCAGTA ACCTAGACTT AGGGCAAGGC GGCACTGACC TAGTTTGAAG AGATTTCCGA 13020

AGAGTATAAA TTTTAATATT TTCTTGTGTT ATTCCTTGAC AATTCAATTT GGAAAATATA 13080

TGATAAAGAT AATGACAGCG GTGTCATTCT ATCTATTTTA AGAAAAGTAA TAATCAATTG 13140

TTAAAAATAG TAAAAAAATT GGAGGTTCTG ATGAAATATT TTGTTCCG 13188

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32768 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

AACGAGTGCA TCAGTCTCAG CAAGCACCAG TGCGTCGGCC TCAGCAAGCA CCAGCGCGTC 60

TGAATCCGCA TCAACCAGTG CCTCAGCTTC AGCAAGTACC TCAGCATCTG AATCAGCATC 120

AACAAGTGCA TCGGCTTCAG CAAGCACAAG TGCTTCAGCC TCAGCAAGTA TCTCAGCGTC 180

TGAATCGGCA TCAACGAGTG CGTCCGCTTC AGCAAGTACT AGCGCCTCAG CATCAGCGTC 240

AACAAGTGCT TCGGCTTCAG CGTCAACGAG TGCGTCTGAG TCAGCATCAA CGAGTACGTC 300

AGCCTCAGCA AGCACATCAG CTTCTGAATC TGCATCAACC AGTGCGTCAG CCTCAGCATC 360

GACAAGCGCC TCAGCTTCAG CAAGTACCAG TGCGTCAGCC TCAGCAAGTA CCAGTGCTTC 420

AGCCTCAGCG TCGACAAGTG CGTCGGCCTC AACCAGTGCA TCTGAATCGG CATCAACCAG 480

TGCGTCAGCC TCAGCAAGTA CTAGCGCCTC AGCCTCAGCA TCAACGAGTG CGTCCGCTTC 540

AGCAAGTACT AGTGCATCAG CATCAGCATC AACGAGTGCA TCGGCTTCAG CAAGTACCAG 600

CGCCTCAGCT TCAGCAAGCA CCAGTGCGTC AGCCTCAGCA AGTACCAGCG CCTCAGCCTC 660

AGCAAGCACC AGTGCCTCAG CTTCAGCAAG TACCAGTGCG TCAGCCTCAG CGTCGACAAG 720

TGCGTCGGCT TCAGCAAGTA CCTCAGCGTC TGAATCAGCA TCAACGAGTG CATCAGCTTC 780

AGCATCAACA AGTGCTTCAG CTTCAGCAAG TATCTCAGCG TCTGAATCGG CATCAACGAG 840

TGCGTCCGCT TCAGCAAGTA CTAGCGCCTC AGCATCAGCG TCAACAAGTG CTTCGGCTTC 900

AGCGTCAACG AGTGCGTCTG AGTCAGCATC AACGAGTACG TCAGCCTCAG CAAGCACATC 960

AGCTTCTGAA TCTGCATCAA CCAGTGCGTC AGCCTCAGCA TCGACAAGCG CCTCAGCTTC 1020

AGCAAGTACC AGTGCGTCAG CCTCAGCAAG TACCAGTGCT TCAGCCTCAG CGTCGACAAG 1080

TGCGTCGGCC TCAACCAGTG CATCTGAATC GGCATCAACC AGTGCGTCAg CCTCAGCAAG 1140

TACTAGCGCC TCAGCCTCAG CATCAACGAG TGCGTCCGCT TCAGCAAGTA CTAGTGCATC 1200

AGCATCAGCA TCAACGAGTG CATCGGCTTC AGCAAGTACC AGCGCCTCAG CTTCAGCAAG 1260

CACCAGTGCG TCAGnCTCAG CAAGTACCAG CGCCTCAGCC TCAGCAAGCA CCAGTGCCTC 1320

AGCTTCAGCA AGTACCAGTG CGTCAgCCTC AGCGTCGACA AGTGCGTCGG CTTCAGCAAG 1380

TACCTCAGCG TCTGAATCAG CATCAACGAG TGCATCAGCT TCAGCATCAA CAAGTGCTTC 1440

AGCTTCAGCA AGTACCAGTG CGTCGGCTTC AGCATCAACG AGTGCTTCAG TCTCAGCGTC 1500

AACCAGTGCC TCTGAATCAG CATCAACAAG TGCCTCGGCT TCAGCAAGCA CCAGTGCGTC 1560

GGCTTCAGCA AGTACTAGTG CATCGGCTTC AGCATCGACA AGTGCGTCTG AATCGGCATC 1620

AACGAGTGCT TCGGCTTCAG CATCAACGAG TGCGTCAGCC TCAGCAAGCA CATCAGCTTC 1680

TGAATCTGCA TCAACCAGTG CGTCCGCTTC AGCGTCAACC AGTGCGTCGG CTTCAGCGTC 1740

GACAAGTGCT TCGGCTTCAG CATCAACGAG TGCGTCGGCC TCAGCAAGCG CAAGTACCTC 1800

AGCGTCAGct TCCGCCTCAA CCAGTGCGTC GGCTTCAGCA AGCACAAGTG CGTCAGCCTC 1860

AGCAAGTATC TCAGCGTCTG AATCGGCATC AACGAGTGCG TCTGAGTCAG CATCAACGAG 1920

TACGTCAGCC TCAGCAAGCA CATCAGCTTC TGAATCTGCA TCAACCAGTG CGTCAGCCTC 1980

AGCATCGACA AGCGCCTCAG CTTCAGCAAG TACCAGTGCT TCAGCCTCAG CGTCGACAAG 2040

TGCGTCGGCC TCAACCAGTG CATCTGAATC GGCATCAACC AGTGCGTCAG CCTCAGCAAG 2100

TACTAGTGCA TCAGCTTCAG CATCAACGAG TGCATCGGCT TCAGCATCAA CCAGTGCCTC 2160

GGCTTCAGCG TCAACCAGTG CGTCAGCTTC AGCAAGTACC AGTGCTTCAG TCTCAGCATC 2220

AACAAGTGCT TCAGCCTCAG CATCGACAAG TGCCTCGGCT TCAGCAAGCA CATCAGCATC 2280

TGAATCAGCG TCAACCAGTG CTTCGGCTTC AGCAAGTACC AGTGCTTCAG CTTCAGCATC 2340

AACCAGCGCC TCGGCCTCAG CAAGCACCTC AGCTTCTGAA TCGGCCTCAA CCAGCGCCTC 2400

GGCCTCAGCA AGCACCTCAG CTTCTGAATC GGCCTCAACC AGCGCCTCAG CCTCAGCATC 2460

AACGAGTGCT TCGGCTTCAG CAAGCACAAG CGCCTCGGGT TCAGCATCAA CGAGTACGTC 2520

AGCTTCAGCG TCAACCAGTG CTTCAGCCTC AGCATCAACA AGTGCGTCAG CCTCAGCAAG 2580

TATCTCAGCG TCTGAATCGG CATCAACGAG TGCGTCTGAG TCAGCATCAA CGAGTACGTC 2640

AGCCTCAGCA AGCACCTCAG CTTCTGAATC GGCCTCAACC AGTGCGTCAG CCTCAGCATC 2700

GACAAGCGCC TCAGCTTCAG CAAGTACCAG TGCTTCAGCC TCAGCGTCGA CAAGTGCGTC 2760

GGCCTCAACC AGTGCATCTG AATCGGCATC AACCAGTGCG TCAGCCTCAG CAAGTACTAG 2820

TGCATCGGCT TCAGCATCAA CCAGTGCCTC GGCTTCAGCG TCAACCAGTG CGTCAGCTTC 2880

AGCAAGTACC AGTGCTTCAG TCTCAGCATC AACAAGTGCT TCAGCCTCAG CATCGACAAG 2940

TGCCTCGGCT TCAGCAAGCA CATCAGCATC TGAATCAGCG TCGACAAGCG CCTCAGCTTC 3000

AGCAAGTACC AGTGCGTCAG CCTCAGCGTC GACAAGTGCG TCAGCCTCAG CAAGTACTAG 3060

TGCATCAGCT TCAGCATCAA CGAGTGCATC GGCTTCGGCG TCAACCAGTG CATCAGAGTC 3120

AGCAAGTACC AGTGCGTCAg CTTCCGCATC AACAAGTGCC TCGGCTTCAG CAAGCACCAG 3180

TGCGTCGGCT TCAGCAAGTA CTAGCGCCTC AGCCTCAGCC TCAACCAGTG CGTCAGCCTC 3240

AGCAAGTATC TCAGCGTCTG AATCGGCATC AACGAGTGCG TCCGCTTCAG CAAGTACTAG 3300

CGCCTCAGCC TCAGCGTCAA CAAGTGCATC GGCTTCAGCG TCAACGAGTG CGTCTGAATC 3360

GGCATCAACG AGTGCGTCCG CTTCAGCAAG TACTAGCGCC TCAGCCTCAG CGTCAACAAG 3420

TGCATCGGCT TCAGCATCAA CGAGTGCGTC CGCTTCAGCA AGTACTAGCG CCTCAGCCTC 3480

AGCGTCAACA AGTGCATCGG CTTCAGCGTC AACGAGTGCG TCTGAGTCAG CATCAACGAG 3540

TGCGTCAGCC TCAGCAAGCA CATCAGCTTC TGAATCTGCA TCAACCAGTG CGTCAGCCTC 3600

AGCATCGACA AGCGCCTCAG CTTCAGCAAG TACCAGTGCG TCAGCCTCAG CGTCGACAAG 3660

TGCGTCGGCT TCAGCAAGTA CCAGTGCGTC AGCCTCAGCA AGTACCAGTG CGTCAGCCTC 3720

AGCGTCGACA AGTGCGTCGG CCTCAACCAG TGCATCTGAA TCGGCATCAA CCAGTGCGTC 3780

AGCCTCAGCA AGTACTAGTG CATCAGCTTC AGCATCAACG AGTGCATCGG CTTCAGCATC 3840

AACCAGTGCA TCAGAGTCAG CAAGTACCAG TGCGTCAGCT TCCGCATCAA CAAGTGCCTC 3900

GGCTTCAGCA AGTACTAGCG CCTCAGCCTC AGCGTCAACA AGTGCTTCAG CTTCCGCGTC 3960

AACCAGCGCC TCGGCCTCAG CAAGTATCTC AGCGTCTGAA TCGGCATCAA CAAGTGCCTC 4020

GGCTTCAGCA TCAACGAGTG CATCAGTCTC AGCAAGCACC AGTGCGTCGG CCTCAGCAAG 4080

CACCAGCGCG TCTGAATCCG CATCAACCAG TGCCTCAGCT TCAGCAAGTA CCTCAGCATC 4140

TGAATCAGCA TCAACAAGTG CCTCGGCTTC AGCAAGCACA AGTGCTTCAG CCTCAGCAAG 4200

TATCTCAGCG TCTGAATCGG CATCAACGAG TGCGTCCGCT TCAGCAAGTA CTAGCGCCTC 4260

AGCATCAGCG TCAACAAGTG CTTCGGCTTC AGCGTCAACG AGTGCGTCTG AGTCAGCATC 4320

AACGAGTACG TCAGCCTCAG CAAGCACATC AGCTTCTGAA TCTGCATCAA CCAGTGCGTC 4380

AGCCTCAGCA TCGACAAGCG CCTCAGCTTC AGCAAGTACC AGTGCGTCAG CCTCAGCAAG 4440

TACCAGTGCT TCAGCCTCAG CGTCGACAAG TGCGTCGGCC TCAACCAGTG CATCTGAATC 4500

GGCATCAACC AGTGCGTCAG CCTCAGCAAG TACTAGCGCC TCAGCCTCAG CATCAACGAG 4560

TGCGTCCGCT TCAGCAAGTA CTAGTGCATC AGCTTCAGCA AGTACTAGCG CCTCAGCCTC 4620

AGCGTCGACA AGCGCCTCAG CTTCAGCAAG TACCAGTGCG TCAGCCTCAG CGTCGACAAG 4680

TGCGTCGGCT TCAGCAAGTA CCTCAGCGTC TGAATCAGCA TCAACAAGTG CGTCGGCTTC 4740

AGCATCAACG AGTGCATCAG CTTCAGCATC AACAAGTGCT TCAGCTTCAG CAAGTACCAG 4800

TGCGTCGGCT TCAGCATCAA CGAGTGCTTC AGTCTCAGCG TCAACCAGTG CCTCTGAATC 4860

CGCATCAACA AGTGCCTCGG CTTCAGCAAG CACCAGTGCT TCGGCTTCAG CGTCAACGAG 4920

TGCGTCTGAG TCAGCATCAA CGAGTGCGTC AGCCTCAGCA AGCACATCAG CTTCTGAATC 4980

TGCATCAACC AGTGCGTCAG CTTCCGCATC AACAAGCGCC TCGGCCTCAG CAAGTACAAG 5040

TGCTTCAGCC TCAGCATCAA CCAGTGCATC AGCTTCAGCC TCAACAAGTG CTTCAGCCTC 5100

AGCGTCAACC AGTGCCTCGG CTTCAGCAAG TACCAGTGCG TCAGCTTCAG CAAGCACAAG 5160

TGCGTCAGCT TCAGCATCAA CCAGTGCTTC GGCTTCGGCA TCAACAAGTG CCTCAGCATC 5220

AGCATCAACG AGTGCGTCAG CCTCAGCAAG TACTAGTGCA TCAGCATCAG CATCAACCAG 5280

TGCATCAGCC TCAGCAAGTA TCTCAGCGTC TGAATCGGCA TCAACGAGTG CATCAGCATC 5340

AGCATCAACG AGTGCATCGG CTTCAGCGTC AACCAGTGCA TCAGTCTCAG CAAGCACCAG 5400

TGCGTCGGCT TCAGCATCAA CGAGTGCCTC AGCCTCAGCA AGTATCTCAG CGTCTGAATC 5460

GGCATCAACG AGTGCGTCAG CCTCAGCAAG TACTAGTGCA TCGGCTTCAG CAAGCACCAG 5520

TGCGTCGGCT TCAGCATCAA CCAGTGCCTC AGCCTCAGCA AGTATCTCAG CGTCTGAATC 5580

GGCATCAACG AGTGCGTCAG CCTCAGCAAG TACTAGTGCA TCAGCmTCAG CATCAACGAG 5640

TGCATCGGCT TCAGCAAGTA CCAGCGCCTC AGCTTCAGCA AGCACCAGTG CGTCAGCCTC 5700

AGCAAGTACC AGCGCCTCAG CCTCAGCAAG CACCAGTGCC TCAGCTTCAG CAAGTACCAG 5760

TGCGTCAGcT CAGCATCAAC AAGTGCTTCA GCTTCGGCCT CAACAAGTGC GTCAGCTTCA 5820

GCATCAACGA GTGCGTCGGC TTCAGCAAGC ACCAGTGCCT CGGCCTCAGC AAGCACCAGT 5880

GCTTCAGCTT CAGCATCAAC AAGTGCGTCA GCTTCAGCAA GTACATCAGT TTCAAATTCA 5940

GCAAACCATT CGAACTCACA AGTTGGAAAT ACTTCTGGAT CGACAGGTAA ATCCCAAAAA 6000

GAATTGCCTA ATACAGGTAC TGAGTCGTCA ATTGGATCTG TGTTACTTGG AGTTCTAGCA 6060

GCTGTTACAG GTATTGGATT GGTTGCGAAA CGCCGTAAAC GTGATGAAGA AGAGTAAGAC 6120

AACCTGTAAA GTTAGGCTAA ACTAACTCGC GCACATAAAT CAAGGAGAAA ATTGCTAGTG 6180

GATGATAAAA TAACAGTCAT TGTACCAGTA TACAATGTGG AAAACTATCT GAGGAAGTGC 6240

CTAGATAGTA TTATTACTCA AACATATAAA AATATTGAGA TTGTTGTCGT TAATGATGGT 6300

TCTACGGATG CTTCAGGTGA AATTTGTAAA GAATTTTCAG AAATGGATCA CCGAATTCTC 6360

TATATAGAAC AAGAAAATGC TGGTCTTTCT GCCGCACGAA ACACCGGTCT GAATAATATG 6420

TCCGGAAATT ATGTGACCTT TGTGGACTCG GATGATTGGA TTGAGCAAGA TTATGTAGAA 6480

ACTCTATATA AAAAAATAGT AGAGTATCAG GCTGATATTG CAGTTGGTAA TTATTATTCT 6540

TTCAACGAAA GTGAAGGAAT GTTCTACTTT CATATATTGG GAGACTCCTA TTATGAGAAA 6600

GTATATGATA ATGTTTCTAT CTTTGAGAAC TTGTATGAAA CTCAAGAAAT GAAGAGTTTT 6660

GCTTTGATAT CTGCTTGGGG TAAACTCTAT AAGGCAAGAT TGTTTGAGCA GTTGCGCTTT 6720

GACATAGGTA AATTAGGAGA AGATGGTTAC CTCAATCAAA AGGTATATTT ATTATCAGAA 6780

AAGGTAATTT ATTTAAATAA AAGTCTTTAT GCTTATCGGA TTAGAAAAGG TAGTTTATCA 6840

AGAGTTTGGA CAGAAAAGTG GATGCACGCT TTAGTTGATG CTATGTCTGA ACGTATTACG 6900

CTACTAGCTA ATATGGGTTA TCCTCTAGAG AAACACTTGG CAGTTTATCG TCAGATGTTG 6960

GAAGTCAGTC TCGCCAACGG TCAAGCTAGT GGTTTATCTG ACACAGCAAC GTATAAAGAG 7020

TTTGAAATGA AACAAAGGCT TTTAAATCAG CTATCGAGAC AAGAGGAAAG TGAAAAGAAA 7080

GCCATTGTCC TCGCAGCAAA CTATGGCTAT GTAGACCAAG TTTTAACGAC AATCAAGTCT 7140

ATTTGTTATC ATAATCGTTC GATTCGTTTT TATCTGATTC ATAGCGATTT TCCAAATGAA 7200

TGGATTAAGC AATTAAATAA GCGCTTAGAG AAGTTTGACT CAGAAATTAT TAATTGTCGG 7260

GTAACTTCTG AGCAAATTTC ATGTTATAAA TCGGATATTA GTTACACAGT CTTTTTACGC 7320

TATTTCATAG CTGATTTCGT GCAAGAAGAC AAGGCCCTCT ACTTGGACTG TGATCTAGTT 7380

GTAACGAAAA ATCTGGATGA CTTGTTTGCT ACAGACTTAC AAGATTATCC TTTGGCTGCT 7440

GTTAGAGATT TTGGGGGCAG AGCTTATTTT GGTCAAGAAA TCTTTAATGC CGGTGTTCTC 7500

TTGGTAAACA ATGCTTTTTG GAAAAAAGAG AATATGACCC AAAAATTAAT TGATGTAACC 7560

AATGAATGGC ATGATAAGGT GGATCAGGCA GATCAGAGCA TCTTGAATAT GCTTTTTGAA 7620

CATAAATGGT TGGAATTGGA CTTTGATTAT AATCATATTG TCATTCATAA ACAGTTTGCT 7680

GATTATCAAT TGCCTGAGGG TCAGGATTAT CCTGCTATTA TTCACTATCT TTCTCATCGG 7740

AAACCGTGGA AAGATTTGGC GGCCCAAACC TATCGTGAAG TTTGGTGGTA CTATCATGGG 7800

CTTGAATGGA CAGAATTGGG ACAAAACCAT CATTTACATC CATTACAAAG ATCTCACATC 7860

TATCCAATAA AGGAACCTTT CACTTGTCTA ATCTATACTG CCTCAGACCA TATTGAACAA 7920

ATTGAGACAT TGGTTCAATC CTTGCCTGAT ATTCAGTTTA AGATAGCAGC TAGAGTAATA 7980

GTTAGTGATC GATTGGCTCA GATGACAATT TATCCAAACG TGACTATATT TAACGGAATT 8040

CACTATTTGG TAGATGTCGA TAATGAATTG GTAGAAACCA GTCAAGTACT TTTAGATATT 8100

AATCATGGCG AAAAGACAGA AGAAATTCTC GATCAATTTG CTAATCTTGG CAAGCCTATC 8160

TTATCCTTTG AAAATACTAA AACCTATGAA GTAGGTCAGG AGGCATATGC TGTTGACCAA 8220

GTTCAAGCAA TGATTGAAAA ATTGAGAGAA ATAAGCAAAT GAAGAAAAAT CATTTAGTAG 8280

GAGATGCTCT GATTTTGACG GTTAGTGATC AGATTGAAGA GTTGGATTAT TTTTTATAAA 8340

ATTTCTCCGT TCATCATATA TGAAAGTTGT TCAAACATCA GAGTGCTTTA TAAAATATAA 8400

ATAGACCTAA AGATATTTAA TATGAACTGC ACCCCAAAAG TTAGACAGAA AAAATCTAAC 8460

TTTTTGGsGT CAGTACAATA TTAGGGTGTG ATTAATTATC TTTTTAGGTG AAAATGATTC 8520

TATATTATAG CTGTTTGATA CGAAATTTAT TATAAGGAAA TTATGTTAAT GAATACAAAA 8580

TCTATAGTTT TTAATGCAGA TAATGATTAT GTAGATAAAT TAGAAACTGC AATTAAATCT 8640

ATTTGTTGTT ATAATAATTG TTTAAAATTT TATGTATTTA ATGATGATAT TGCGTCAGAG 8700

TGGTTTTTGA TGATGAATAA GCGATTGAAG ACTATACAAT CTGAAATCGT TAATGTAAAG 8760

ATTGTAGATC ATGTTCTTAA AAAGTTTCAT TTACCGTTAA AGAATTTAAG TTATGCCACT 8820

580

TTCTTTCGTT ATTTTATACC TAATTTTGTC AAAGAAAGTC GTGCTTTATA CCTAGATTCT 8880

GACATCATTG TTACAGGAAG TTTAGACTAT TTATTTGATA TAGAACTAGA TGGTTATGCC 8940

TTGGCAGCAG TAGAAGATTC TTTTGGTGAT GTTCCTTCTA CCAATTTTAA CTCCGGAATG 9000

TTATTAGTTA ATGTAGATAC TTGGAGAGAT GAAGATGCTT GTTCGAAACT GTTAGAACTG 9060

ACCAATCAAT ATCATGAAAC AGCATATGGA GATCAAGGAA TTTTAAATAT GTTATTCCAT 9120

GATAGATGGA AAAGATTAGA CCGAAATTTT AATTTTATGG TGGGGATGGA TAGCGTCGCA 9180

CACATAGAAG GAAATCATAA ATGGTATGAG ATTTCTGAGT TGAAAAATGG AGATTTACCT 9240

AGTGTTATAC ATTATACTGG GGTAAAACCT TGGGAAATAA TTTCCAATAA TCGCTTTAGA 9300

GAAGTTTGGT GGTTTTATAA TCTGTTAGAA TGGTCTGATA TTTTATTGAG AAAAGACATT 9360

ATTAGTCGTA GTTTCGAAGA ACTTGTATAC AGTCCTAAAG CTCATACAGC AATTTTTACA 9420

GCTAGTTGTG AGATGGAGCA TGTAGAATAT TTGATAGAAA ATTTACCAGA GGTACATTTT 9480

TCTATACTAG CACATACATA TTTTGCGTCT AGTGTCGTTG CTTTATTAAG ATATAGCAAT 9540

GTTACGATTT ATCCTTGTTT TTCTCCATTT GATTATCGAA AAATTTTGGA TAATTTAGAT 9600

TTTTATTTAG ATATTAATCA TTATAAAGAA GTGGATAATA TTGTATCCGT TGTTCAACAA 9660

CTATCTAAAC CAATTTTTAC CTTTGAAAAT ACTAGTCATG ATATAGGCAA TCAAACTAAT 9720

ATATTTTCTT CAACCGAACC AAACAAAATG GTAGAGGCTA TTAGACAATT TATAGGAGAA 9780

TAAGTTTATG GCAGACGAAC TAATTAGTAT TGTAGTTCCA ATCTACAACG TTGAGAATTA 9840

TTTGCGAATG TGTTTGGATA GCATTCAGAA TCAGACGTAT CAAAATTTTG AGTGTTTATT 9900

AATCAATGAT GGCTCTCCAG ATCATTCATC CAAAATATGT GAAGAATTTG TAGAGAAAGA 9960

TTCTCGTTTC AAATATTTTG AGAAAGCAAA CGGCGGTCTT TCATCAGCTC GTAACCTAGG 10020

TATTGAATGT TCGGGGGGGG GCGTACATTA CTTTTGTAGA CTCTGATGAT TGGTTGGAAC 10080

ATGATGCTTT AGACCGATTA TATGGTGCTT TGAAAAAGGA AAACGCAGAT ATTAGTATCG 10140

GGCGTTATAA TTCTTATGAT GAAACACGCT ATGTGTATAT GACTTATGTT ACGGATCCAG 10200

ATGATTCTCT AGAAGTGATA GAAGGTAAAG CAATTATGGA TAGGGAAGGT GTCGAAGAAG 10260

TCAGAAATGG GAACTGGACT GTAGCTGTCT TGAAGTTATT CAAGAGAGAG TTACTACAAG 10320

ATTTACCATT TCCTATAGGA AAAATTGCAG AGGATACTTA CTGGACATGG AAGGTACTTC 10380

TAAGAGCTTC GAGGATAGTC TATTTGAATC GTTGTGTTTA CTGGTACCGT GTTGGTTTAT 10440

CTGATACTTT ATCGAATACA TGGAGTGAAA AGCGTATGTA TGATGAAATT GGGGCTAGGG 10500

AAGAAAAGAT AGCTATTTTA GCAAGTTCAG ACTATGACTT GACCAATCAT ATTTTGATTT 10560

ATAAAAATAG ATTACAAAGA GTGATAGCAA AATTAGAAGA ACAAAATATG CAGTTCACAG 10620

AGATTTACAG AAGAATGATG GAAAAATTGT CTTTACTTCC GTAGATAGTA ATAAAAAATG 10680

AGATAGCGTA ATATGAAACT ACATTTAACA AATTTATACG GCATGGCTGG TGATAGTACG 10740

GTTATCTTAG CTCAAAATGC TGTTCAAAAG ATAGCTAGTC AACTGGGATT TAGAGAGGTT 10800

GGTATTTATT TTTACAACAT TGCTTCAGAT AGTCCTTCTG AAATGAATAA GCGTCTGGAT 10860

GGTATTATGG CCAGTATCTC TATTGGGGAT ATTTTAGTCT TTCAGTCTCC AACCTGGAAT 10920

GGTTTTGAAT TTGATCGTCT CTTGTTTGAT AAGCTAAAGG ATATGCAGGT GAAAATTATT 10980

TGCTTTATCC ATGATGTTGT TCCCCTCATG TTTGATAGTA ACTATTATCT CATGAAAGAT 11040

TATCTGTATA TGTATAATCT ATCAGATGTT TTGATAGTGC CGTCAGAGAG AATGAAAACA 11100

CGCCTGATGG AAGAAGGATT GACGACTAAG AAGATTCTTG TTCAAGGGAT GTGGGATCAT 11160

CCTCATGATT TATCCTTATA CACCCCTGCT TTTAAAAAAG AACTTTTTTT TGCTGGAAGT 11220

TTAGAGCGTT TTCCAGACTT ACAAAATTGG TCTCAAGATA CGCCTTTGAG AGTATTTTCA 11280

AATAAAGGGG AAGCTAGTTC TAGTGCTAGA AGTCTCAGCA TCGAAGGATG GAAAAAAGAT 11340

GAGGAATTGT TGCTAGAATT ATCAAAGGGT GGATTTGGCC TTGTCTGGGG AACCcATCAA 11400

AATGAGGGAG AAAGTAACCA ATACTATACC TTGAATATAT CTCATAAGGT GAGTACCTAT 11460

CTAACAGCGG GCATTCCAGT CATTGTACCA AGTAGCTTGT CAACTGCTAA ATTTATAGTA 11520

GATCAAGGCT TGGGCTTTAT GGCGGATAGT CTGGAAGAGG TTCATGAGAT AGTTGATAAA 11580

ATGAATCTAC AAGAATATCA AGAAATGACG AATCGTATCA AGACCTTTAG CTATTTGTTA 11640

AAAGAGGGCT ATTTCACTAA AAAGTTATTG GTAGATGCAA TCTATCACTT GGGAATTGAT 11700

TAAGGGAATG AAATGAACAA AACAATTGTA CTAGCAGGGG ATCGCAATTA CACCAGGCAG 11760

TTAGAAACAA CGATAAAATC TATTTTATAC CACAATCGAG ATGTTAAGAT TTATATTTTG 11820

AATCAAGATA TCATGCCAGA TTGGTTTCGC AAACCACGAA AAATAGCTCG CATGTTAGGT 11880

AGTGAGATTA TCGATGTTAA ACTACCTGAA CAAACTGTGT TTCAAGATTG GGAAAAGCAA 11940

GATCACATTA GTAGCATTAC TTATGCTAGA TATTTTATTG CAGATTATAT CCAAGAAGAT 12000

AAGGTTTTAT ATTTAGACAG TGATTTGATT GTAAATACTT CTTTAGAGAA ATTATTTAGT 12060

ATTTGTTTAG AAGAAAAATC ACTCGCAGCA GTTAAAGATA CAGATGGAAT TACATTTAAT 12120

GCAGGTGTTT TATTAATCAA CAATAAAAAA TGGCGTCAAG AGAAATTAAA AGAACGACTA 12180

ATTGAACAGA GCATTGTTAC AATGAAGGAA GTTGAAGAAG GCCGTTTCGA GCATTTTAAT 12240

GGTGATCAAA CGATTTTTAA TCAGGTCTTG CAAGATGATT GGTTAGAACT AGGTCGAGCT 12300

TATAATTTAC AAGTAGGGCA TGATATTGTG GCTTTGTATA ACAATTGGCA GGAACATCTG 12360

GCTTTTAATG ATAAACCAGT GGTGATTCAT TTTACGACCT ACAGAAAACC CTGGACTACC 12420

TTGACAGCCA ATCGTTATCG TGATTTATGG TGGGAATTCC ATGATTTGGA GTGGAGTCAG 12480

ATTTTACAAC ACCATATGGG AGAATTTGAA CTAATATCGC CTCTAGATAA GGAATTTTCT 12540

TGCTTAACCT TAACGAATTC CCAAGATTTA GAAGGAATAG AAGAGCTAGT TACAGCTCTA 12600

CCTGAGGTGG TATTTCATAT CGCAGCTTGG ACGGATATGG GAGATAAATT AAAAAAATTA 12660

GCTGTATATA ATAATGTGAG ATTGCATCCA CAAATTGTTC CACCGGTCTT AGATAAGCTG 12720

AAAAAGTCAA CAAATCTATA TTTGGATATC AATCATGGTA GTGCAGATGA GAACTTTTTA 12780

AAATCTTTGC AAGAACAAGA AAAAACGCTA CTAGCTTTTC AATCGACTCA GCACGGAGAG 12840

TTAGGACAAA TCGTTTTCGA AAATGGGAAA GTTTCCTTTA TGATTGATAC GATTAAAGAT 12900

TTTAAGAAAA ACGGACATCT TACCTGTTTT CGACAACTTC CAAGTTTAAC TTGTTTAACG 12960

TTTACGGCTT CTCAGTATAT CGAACAATTG GATTACTTGG CTGGACAGTT GCCAAATGTT 13020

GTTTTTCAAA TTGCTGCTTG GACAGCTATG GGGCCAAAAT TATATGATTT GTCTAATCGT 13080

TATCCTAATA TTCAGCTCTA TCCGGCAATT TCTAGAGATA AGCTAGACGA GTTGAAGGAG 13140

AAGATGGATG CTTATTTAGA TATCAACCTA CTGACTTCAA CATCCGATAT CGTTGCAGAA 13200

ATGGCTCATC TATCTAAACC TATACTAGCC TTTTATAAAT CTCAAAATGG GAATAATGGC 13260

CAAAGGTTGT ATTCAAGTGA ACATCCTGAA CGAATGTTGG CTGATTTGCA AAAATTGATA 13320

ACTAAGGATA TGCTAGAAAA ACCGCTTGAT ATAATCCAGG TGAAAGGGAT AGATGAAACC 13380

TTGGATTATA TTATTGAACA CAACTCTTCT TTAGTTCGTT TTGGAGATGG GGAAATCAAT 13440

ATGCTTGCAG GGCATTCAAT TCCCTACCAG GATTATGATG AAGAGTTGGT TTCAATCATG 13500

AGGGACATTA TCGGCCAAGA AAGTCGAGAA GATTTAGTAG TGTGCCTTCC TGATGCTTTT 13560

ACAGATCGTT TTAGGTTTAC ATCGTGGGCG ATTCCATTTT GGAAAGATCA CATGGATCAT 13620

TATATGGATT TTTACAGAGA GTTATGCAGT GATTCATGGT ATGGCTCAAC CTTTGTATCT 13680

CGCCCTTATA TCGATTTTGA AGACAAGAGT CAAGCTAAAG CTCAATTTGA AAAATTGAAA 13740

AGCATTTGGG AAAACCGTGA CTTACTGATA GTCGAAGGTG CGACTTCTCG TTCAGGTGTC 13800

GGAAATGATT TATTCGATGA GGCAAATTCT ATTAAGCGAA TTATCTGTCC TTCTCATAGT 13860

GCCTTTTCTA GAGTTCATGA ACTTGAACAA GAAATTGAAA AGTATGCTGG TGGTCGCTTG 13920

ATTTTATGTA TGCTTGGACC TACAGCAAAA GTTCTGAGTT ATAATCTATG CCAGATGGGC 13980

TATCAAGTTT TGGATGTAGG CCATATTGAC TCAGAGTATG AATGGATGAA AATGGGAGCT 14040

AAAACTAAGG TTAAATTTTC TCATAAACAT ACTGCAGAAC ATAATTTCGA CCAAGATATT 14100

GAATTTATTG ATGATGAAAC CTATAACAGT CAGATTGTTG CACGAATATT AAACTAGACT 14160

ATTTAAAATA AATGATAAGG ATTTAAAATG AGAAATACCA AACGCGCTGT AGTATTTGCA 14220

GGTGATTACG CTTATATTCG ACAAATCGAA ACGGCGATGA AGTCACTCTG TAGACACAAT 14280

AGTCATTTGA AAATTTATCT GCTAAATCAG GACATTCCTC AGGAATGGTT TAGTCAAATA 14340

AGAATATATT TACAAGAGAT GGGGGGCGAC TTGATTGACT GCAAGTTAAT TGGCTCACAG 14400

TTTCAAATGA ATTGGTCTAA TAAATTACCT CATATCAATC ATATGACATT TGCACGCTAT 14460

TTTATTCCAG ATTTTGTAAC AGAAGATAAA GTTCTCTATC TAGATAGTGA TTTGATTGTG 14520

ACTGGTGATT TGACCGATTT GTTTGAATTA GACTTAGGTG AAAATTATTT GGCAGCAGCT 14580

CGTTCTTGCT TTGGAGCAGG AGTCGGCTTC AATGCTGGTG TTCTCTTGAT TAACAACAAA 14640

AAATGGGGAT CTGAAACTAT TCGACAAAAA TTGATTGACT TAACAGAAAA AGAACATGAG 14700

AATGTGGAAG AAGGAGACCA GTCAATTTTG AATATGTTGT TTAAAGATCA ATATAGTTCC 14760

CTTGAAGATC AATATAATTT TCAAATAGGA TATGATTATG GGGCGGCAAC CTTTAAACAT 14820

CAATTCATTT TTGATATTCC GCTCGAACCA CTGCCACTAA TTTTACACTA TATTTCTCAG 14880

GATAAGCCTT GGAATCAATT TTCTGTTGGA CGTCTAAGAG AAGTTTGGTG GGAATACTCT 14940

TTGATGGATT GGTCTGTTAT TTTAAATGAA TGGTTTTCAA AGAGTGTGAA GTACCCTAGT 15000

AAATCACAAA TATTTAAGTT GCAATGTGTT AATTTAACGA ATTCTTGGTG TGTCGAGAAA 15060

ATCGATTATT TGGCGGAGCA ATTGCCAGAA GTTCATTTTC ATATTGTTGC TTATACAAAT 15120

ATGGCAAATG AACTACTAGC TTTAACGCGT TTTCCTAATG TTACCGTATA TCCAAATTCC 15180

TTACCAATGT TATTGGAACA AATAGTAATA GCTTCAGATT TGTATTTGGA TTTGAATCAT 15240

GATCGAAAAT TAGAAGATGC ATATGAGTTT GTGCTTAAGT ACAAAAAACC AATGATAGCT 15300

TTCGACAATA CTTGCTCTGA AAATCTTTCT GAGATTTCAT ATGAAGGTAT CTATCCAAGC 15360

TCCATTCCGA AAAAAATGGT TGCAGCAATC AGATCTTACA TGAGGTAGAG AACAGTATGA 15420

GAAAATCAAT AGTATTAGCG GCAGATAATG CCTATCTTAT TCCTTTAGAG ACGACTATAA 15480

AGTCTGTATT GTATCACAAT AGAGATGTTG ATTTTTATAT TCTCAACAGT GATATAGCTC 15540

CTGAATGGTT TAAATTATTG GGGAGAAAAA TGGAAGTTGT GAATTCTACA ATTCGCAGTG 15600

TACACATTGA TAAAGAACTT TTTGAAAGCT ATAAAACAGG ACCTCATATA AATTATGCTT 15660

CTTACTTTAG ATTTTTTGCG ACAGAAGTGG TTGAATCTGA TAGGGTATTG TATCTGGATT 15720

CCGATATCAT TGTAACTGGG GAACTAGCTA CTTTGTTTGA GATAGATCTC AAAGGATATT 15780

CAATTGGTGC TGTTGATGAT GTCTATGCCT ATGAAGGACG AAAATCTGGA TTTAATACTG 15840

GTATGTTACT AATGGATGTT GCAAAGTGGA AAGAACATTC TATTGTCAAT AGTTTATTGG 15900

AATTAGCGGC CGAGCAGAAT CAAGTTGTTC ATCTTGGGGA TCAGAGTATT TTAAATATTT 15960

ATTTTGAGGA TAATTGGCTA GCCTTAGATA AAACATATAA TTATATGGTG GGTATTGATA 16020

TTTATCACCT TGCTCAAGAA TGTGAACGTC TAGATGACAA TCCACCTACA ATTGTTCACT 16080

ATGCTAGTCA TGATAAACCT TGGAATACAT ATAGTATATC TAGACTACGT GAATTATGGT 16140

GGGTTTATAG AGATTTGGAT TGGTCAGAGA TTGCTTTTCA ACGTTCCGAT TTAAATTATT 16200

TTGAAAGAAG CAATCAGTCT AAAAAACAAG TGATGCTTGT GACATGGAGT GCAGATATAA 16260

AACATTTAGA GTATTTAGTA CAACGGTTAC CTGATTGGCA TTTTCATTTG GCTGCACCGT 16320

GTGATTGTTC TGAGGAGCTG ACCTCTCTAT CACAGTATAC GAATGTAACA GTATATCAAA 16380

ATGTATTACA TAGTAGAATT GATTGGCTAT TGGACGATTC TATAGTTTAT TTAGATATTA 16440

ATACAGGTGG AGAGGTTTTT AATGTAGTTA CAAGGGCACA AGAAAGTGGC AAGAAAATCT 16500

TCGCTTTTGA TATCACACGT AAAAGTATGG ATGATGGACT CTATGACGGT ATTTTTTCTG 16560

TGGAGAGACC AGATGATTTA GTGGATAGAA TGAAGAATAT AGAGATAGAG TAATGAGTGA 16620

ATTAATTAGT GTTGTGGTAC CGATATACAA TACGGGAAAA TATTTAGTGG AGTGTGTCGA 16680

GCATATTCTG AAGCAAACCT ATCAAAATAT AGAAATTATT TTAGTTGATG ACGGTTCTAC 16740

GGATAATTCT GGGGAAATTT GTGATGCTTT TATGATGCAA GATAATCGTG TGCGAGTATT 16800

GCATCAAGAA AATAAGGGGG GGGCAGCACA AGCTAAAAAT ATGGGGATTA GTGTAGCTAA 16860

GGGAGAGTAC ATCACGATTG TTGATTCAGA TGATATCGTA AAAGAAAATA TGATTGAAAC 16920

TCTTTATCAG CAAGTCCAAG AAAAGGATGC AGATGTTGTT ATAGGGAATT ACTATAATTA 16980

TGACGAAAGT GACGGGAATT TTTATTTTTA TGTAACAGGG CAAGATTTTT GCGTCGAAGA 17040

ATTAGCTATA CAAGAAATTA TGAACCGTCA AGCAGGAGAT TGGAAATTCA ATAGCTCGGC 17100

CTTTATATTG CCGACATTTA AGTTGATTAA AAAAGAATTA TTCAATGAAG TTCACTTTTC 17160

AAATGGTCGC CGCTTTGATG ATGAAGCAAC TATGCATCGC TTTTATCTTT TAGCCTCTAA 17220

AATCGTCTTT ATAAACGATA ATCTCTATCT GTATAGAAGA CGTTCAGGAA GCATCATGAG 17280

AACGGAATTT GATCTTTCCT GGGCAAGAGA TATTGTTGAA GTGTTTTCTA AGAAAATATC 17340

GGATTGTGTC TTGGCTGGTT TGGATGTCTC CGTTCTGCGT ATTCGATTTG TCAATCTTTT 17400

AAAAGATTAT AAGCAAACTT TAGAATACCA TCAATTAACA GATACTGAGG AATATAAAGA 17460

TATTTGTTTC AGATTAAAGT TGTTTTTTGA TGCAGAACAA AGAAATGGTA AAAGTTGAAA 17520

TAAAAGAATT GTTATTTACC ATATCACAAA CAATGAAGGT GAGGGGAGTG TTTTATGACT 17580

AAGATTTATT CGTCAATAGC AGTAAAAAAA GGACTATTTA CCTCATTTCT ACTGTTTATC 17640

TATGTATTGG GAAGTCGTAT TATTCTCCCT TTTGTTGACC TAAATACTAA AGATTTTTTA 17700

GGAGGTTCAA CAGCCTATCT AGCCTTCTCA GCCGCCCTAA CAGGTGGGAA TCTAAGAAGT 17760

TTATCAATTT TTTCTGTTGG ATTATCCCCT TGGATGTCCG CCATGATTTT ATGGCAGATG 17820

TTTTCTTTTT CTAAACGGTT GGGTTTAACA TCTACGTCTA TAGAAATACA AGATCGCCGT 17880

AAAATGTACC TGACCTTGCT AATTGCTGTG ATTCAATCCT TGGCAGTTAG CTTGAGACTG 17940

CCAGTACAAT CCTCCTATTC TGCAATATTG GTTGTTCTAA TGAATACAAT ATTGCTGATA 18000

GCAGGAACAT TTTTTCTTGT TTGGTTGTCA GATTTAAATG CGAGTATGGG GATTGGAGGT 18060

TCTATTGTAA TCCTCCTATC CAGTATGGTT TTAAATATTC CTCAGGATGT TTTGGAAACA 18120

TTTCAGACAG TACACATTCC AACAGGGATT ATTGTGTTAC TTGCTTTATT AACCCTTGTC 18180

TTTTCTTATT TACTTGCCCT TATGTATCGA GCTCGCTATT TGGTTCCTGT TAATAAAATT 18240

GGCTTACACA ATCGATTTAA ACGCTATTCT TATCTCGAAA TCATGTTGAA TCCTGCAGGT 18300

GGGATGCCTT ATATGTATGT GATGAGTTTT CTTAGTGTAC CAGCTTATTT GTTCATCTTG 18360

TTGGGATTTA TTTTCCCTAA TCATTCAGGG TTAGCGGCTT TATCAAAGGA ATTTATGGTT 18420

GGAAAGCCTT TGTGGGTCTA TGTTTATATT TCGGTCTTAT TTTTATTTAG TATCATTTTT 18480

GCTTTTGTTA CGATGAATGG AGAAGAGATT GCAGACCGTA TGAAAAAATC TGGAGAATAC 18540

ATTTATGGTA TTTATCCAGG TGCGGATACT AGTCGATTTA TTAATCGATT GGTCCTTCGT 18600

TTCTCAGTCA TAGGTGGTCT CTTTAATGTG ATTATGGCAG GTGGTCCCAT GCTTTTTGTT 18660

TTGTTTGATG AAAAGTTATT ACGATTGGCA ATGATTCCAG GCTTATTTAT GATGTTCGGG 18720

GGCATGATTT TTACGATTAG AGACGAGGTC AAGGCTTTAA GGCTAAATGA GACCTATAGA 18780

CCTTTGATTT AGGAGACTTT TATGTATTAT TTTATTCCAG CTTGGTATGG GTCAGAAAGA 18840

ACATGGCATG CAGATATCAC TCCATGGTAT TTTTCTCATT TTCGTCTAGA GTTTGATGAT 18900

ACCTTTCACC AGATTCGGCT CTTTCAAGAG CAAGATATAG ATTCTCGTCT ATTAGTATTA 18960

GCTTACCAGC CTCATCTACG TTATTTTTTA TATAGACATG GTGTGTTAGA AATGGATACT 19020

TATTCCGTTT TTGATGTTAT GCAAGATTTT CATAATCTCC ATACCCAAGT TTTAAGCATT 19080

AGAGATATTG AGTGGGATGA TGACTGTGAA TTTATTTATA GTCCCTTTAC GATTATCGTT 19140

CAAAAAAATG GGAAGAAATT TGCTAAGGTT GAACATGGAG TTGAAGGCTT CATCAGTGAT 19200

ATACAGTATT TTGAACCAAA TGGTCAAATA CATATGCACC ATATCGTGGA TGATCGTGGG 19260

TTTGTATCGA GCATTATCTT TTTTGAAGAT GGGCAAGCAG CCTATCAAGA ATATCTGAAC 19320

CTCAAGGGAG AGTGGCAATT CAGAGAGCGT TTAAAAGAAG GAGGACAGGT AGAAGTCAAT 19380

CCAATTTTGG GTTATCGCTT TAAAATGCTT ACCTATCAAA ATATGGGAGA TCTGGTGGCA 19440

GAATTTTTTG AGAATTATCT GCAAACGTAT GTGAAGGATC AGGATATTTT TATGCTTCCT 19500

TCTCATTCTC ATCATGACCA GTTGGTACTA GATCGTTTAC CTAGTACTAA TCCTAAACTG 19560

TTGAGTCTGT TCATTGGACG TAATCCTCAA GATACCTTTA GGGATTTAGA TGTAACTTTT 19620

GAAAAATCGG ATTTGATTTT GGTGGATAGA GAGGATAGTT TACGATTGTT GCAGGAGTTG 19680

TATCCTGAAC GAATGCATCA ATGTTATCAT TTATCATCTT TTGACACCCG ATTACGATTG 19740

GGACGAAGCC AAACTAAGAA AGAATCCATC ATTTATTTTC AACTGGATTT TGAGCAGGGG 19800

ATTGATAATC AAGCTCTGCT TCAAGTCTTG TCCTTTGTCG CTGAAAATAA GGATACTGAG 19860

GTGATTTTTG GAGCCTTTGC TGCTAGTCAG GAGCAAATGA ATGAGGTTGA AGGGATTGTT 19920

GAGTCTTTCA TCCAAGAAAA CATTCAATCC GAAAATCTGG GAAAGGCGAT TGATTATGGT 19980

GATGCAGAAA ATCCTCTGGA AGAAAATCAA CACCAGGACT TACGCTTACA GTTTGTTAAC 20040

TTGAATGATG AGTTAGATTT GATAAAAACA CTAGAATTTG TCCGTTTGAT TGTGGATTTA 20100

AATAGACATC CTCATCTCTA CACACAGATT GCTGGGATTA GTGCAGGAAT TCCTCAAATC 20160

AACCTAGTTG AAACCGTCTA TGTTGAACAT TTAAAAAATG GTTATTTGTT AGCAGATGTT 20220

ACAGAATTTT CTAAGGCTGC ACATTATTAC ACAGATAGGT TGAAGGAGTG GAATGAGTCC 20280

TTGATATATT CAATTGATAA GATTAAGGAG CACACAGGAC AACAATTTCT TGGAAAATTA 20340

GAGAAATGGA TAGAGGAGGT TAAAAATGTC AAAGGAACTT AATATTTTAC AGATAGGACT 20400

TGCCAATTGG GAAAATCACT ATGACATACC TGAAAATATG AGTTGGTATT ATTTTTACCC 20460

AAACTCATCA AAAGCCCTTC GTGAAATAAT TGAAAAAGAG GATATTAACC GTTTTCATGC 20520

AGTTTTAATA GAAGATGGTC AGTATTCCAG AGACTTATTT TCCTATGTAA AATATTTTGA 20580

ACCTTATACT TTATTTTATA ACCAGAATCT ACAAATAAAT GATAGAGAGG TTGTGGATTT 20640

TCTAAAAAAA CGATGTGCAC AAGCAATTGA TTTTTTAAGT CCCCAACAAC TAATCAATGA 20700

TTTAAGTAAA TCTCTTTTTG GCGGTGGGTA TGGTGATAAA CTCTTTCCTC CGACGATACA 20760

AGTCAATCCA AATTTTACAG GAGCTATTTC TTATCAAGGA TTGGATTATG TCAGTTTGGA 20820

AGGTGAGTTT GGGCAAGATT TTGCCCAGCT TGCCTATTGG GCTTATAATA TTATGGTGCA 20880

AAAAACACTC CCTATTGAGT TGTGGCTTGA ATATGAGAAG GAAGGCAATT GTGACTTTCG 20940

TTTAGTAATC CGTAAAATGT GGAGTGGGTC TGTTGATGAT TTCTTTGAAG AAGTAATAGT 21000

ATCTGAAAAA GACTTGGAGC AAGCACTTTT TATGGATAGT CGAGACGGAG ACTACTTTCT 21060

CTCGATATCT GTTGAAGCAA GAGGTCGTGG AACTATCAAA CTAGGTAATC TTCACCAACG 21120

ATGGAGTCGA AAACAATTTG GTAAGTTTGT ACTTGGTGGA AATATCCTAC ATGATTCCAA 21180

GCGTGATGAA ATAAACTATT TCTTCCATCC AGGTGATTTT AAACCGCCTT TGACTGTCTA 21240

TTTTGCAGGT TATCGACCTG CAGAAGGATT CGAGGGTTAC TTTATGATGA AAACTCTTGG 21300

ATGTCCCTTC ATTTTATTTT CTGATCCACG TTTAGAGGGG GGAGCTTTTT ATCTCGGAAC 21360

GGATGAGCTA GAGGGAAAAG TAAAGGATAC GATCACTCAC TATCTTGATT ATTTAGGCTT 21420

TGATCATAAG GATTTGATTT TATCAGGTCT TTCTATGGGA ACGTTTCCGG CTCTCTATTA 21480

TGGTGCTTCT TTTGAACCCC ATGCCATCAT AGTTGGTAAG CCCTTGGCTA ATTTAGGAAC 21540

TATAGCTAGT CGTGGACGTT TGGACGCACC GGGTGTCTCT AACTTAGCTT TTGATTGTTT 21600

AATTCATCAT ACAGGTGGGA CAAGTTCTCA AGATATGACG GAGTTGGATC AGCGTTTTTG 21660

GAAAATTTTT AAACAAGCAA ATTTTTCAAA GACAACCTTT GGTTTATCCT ATATGAAAGA 21720

TGAAGAAATG GATCCACAAG CCTATGAACA ATTAGTGTCT TATCTGTGTA ATACAGGTGC 21780

GAAGATTTTA TCTAAAGGAA CTGCTGGACG ACACAATGAT GATACAGATA CCAATATTTC 21840

TTGGTTTTTG CACTTTTATA GAATGGTCTT AGAGACTGGT TTTGGAAGGG AGAAAAGATG 21900

ATTATTACAC AGAGACAGTC TATTCATTGG GGAGAAGTTG GTGGGACTTA TATGTATGGA 21960

ACAACTGTAT CTTATTACCC TGACAAAAGT GTTCGTCTGT ATAATCCTCT ATTGCCATCT 22020

GGTGAGATTC TAAAGACTTG GTTTTCTAGT GTCAATTACC AGGCTGCACG AACCCAACCT 22080

CAGCTTCCCT TATTAAAAAG AAAGCAGGAG TATCAACTAT CACTGGTTTT TGACTGTCAG 22140

CCTGAAAATG GAGTTTATAC CAAGATAACT TTTTTTGACC GCTATGGTGA TATTTTAGAA 22200

AAAAAGGTAG AAAAAGTGAA AGATTTCATA TTTACTTATC CAGAAGATAG TTATACTTAT 22260

CGAGTTTCTC TTTTAAGTGC TGGATTTGAG TCCTTAACTT TTTATCATTT TTCTATCAAG 22320

GAGATCAGAA GTGTTTAGAC GTTTAGGTCA AGATTTCCAG CTTAGGAAAG TGAAAAAGAT 22380

TTTAAAGCAG ATTAATGCCC TGAAAGGCAA GATGTCCTCT CTTTCGGATC AAGAATTAGT 22440

AGCTAAAACA GTAGAGTTTC GTCAGCGTCT TTCCGAGGGA GAAAGTCTAG ACGATATTTT 22500

GGTTGAAGCT TTTGCTGTGG TGCGTGAAGC AGATAAGCGG ATTTTAGGGA TGTTTCCTTA 22560

TGATGTTCAA GTCATGGGAG CTATTGTCAT GCACTATGGA AATGTTGCTG AGATGAATAC 22620

GGGGGAAGGT AAGACCTTGA CAGCTACCAT GCCTGTCTAT TTGAACGCTT TTTCAGGAGA 22680

AGGAGTGATG GTTGTGACTC CTAATGAGTA TTTATCAAAG CGTGATGCCG AGGAAATGGG 22740

TCAAGTTTAT CGTTTTCTAG GATTGACCAT TGGTGTACCA TTTACGGAAG ATCCAAAGAA 22800

GGAGATGAAA GCTGAAGAAA AGAAGCTTAT CTATGCTTCG GATATCATCT ACACAACCAA 22860

TAGTAATTTA GGTTTTGATT ATCTAAATGA TAACCTAGCC TCGAATGAAG AAGGTAAGTT 22920

TTTACGACCG TTTAACTATG TGATTATTGA TGAAATTGAT GATATCTTGC TTGATAGTGC 22980

ACAAACTCCT CTGATTATTG CGGGTTCTCC TCGTGTTCAG TCTAATTACT ATGCGATCAT 23040

TGATACACTT GTAACAACCT TGGTCGAAGG AGAGGATTAT ATCTTTAAAG AGGAGAAAGA 23100

GGAGGTTTGG CTCACTACTA AGGGGGCCAA GTCTGCTGAG AATTTCCTAG GGATTGATAA 23160

TTTATACAAG GAAGAGCATG CGTCTTTTGC TCGTCATTTG GTTTATGCGA TTCGAGCTCA 23220

TAAGCTCTTT ACTAAAGATA AGGACTATAT CATTCGTGGA AATGAGATGG TACTGGTTGA 23280

TAAGGGAACA GGGCGTCTAA TGGAAATGAC TAAACTTCAA GGAGGTCTCC ATCAGGCTAT 23340

TGAAGCCAAG GAACATGTCA AATTATCTCC TGAGACGCGG GCTATGGCCT CGATCACCTA 23400

TCAGAGTCTT TTTAAGATGT TTAATAAGAT ATCTGGTATG ACAGGGACAG GTAAGGTCGC 23460

GGAAAAAGAG TTTATTGAAA CTTACAATAT GTCTGTAGTA CGCATTCCAA CCAATCGTCC 23520

GAGACAACGG ATTGACTATC CAGATAATCT ATATATCACT TTACCTGAAA AAGTGTATGC 23580

ATCCTTGGAG TACATCAAGC AATACCATGC TAAGGGAAAT CCTTTACTCG TTTTTGTAGG 23640

CTCAGTTGAA ATGTCTCAAC TCTATTCGTC TCTCTTGTTT CGTGAAGGGA TTGCCCATAA 23700

TGTCCTAAAT GCTAATAATG CGGCGCGTGA GGCTCAGATT ATCTCCGAGT CAGGTCAGAT 23760

GGGGGCTGTG ACAGTGGCTA CCTCTATGGC AGGACGTGGT ACGGATATCA AGCTTGGTAA 23820

AGGAGTCGCA GAGCTTGGGG GCTTGATTGT TATTGGGACT GAGCGGATGG AAAGTCAGCG 23880

GATCGACCTA CAAATTCGTG GCCGTTCTGG TCGTCAGGGA GATCCTGGTA TGAGTAAATT 23940

TTTTGTATCC TTAGAGGATG ATGTTATCAA GAAATTTGGT CCATCTTGGG TGCATAAAAA 24000

GTACAAAGAC TATCAGGTTC AAGATATGAC TCAACCGGAA GTATTGAAAG GTCGTAAATA 24060

CCGGAAACTA GTCGAAAAGG CTCAGCATGC CAGTGATAGT GCTGGACGTT CAGCACGTCG 24120

TCAGACTCTG GAGTATGCTG AAAGTATGAA TATACAACGG GATATAGTCT ATAAAGAGAG 24180

AAATCGTCTA ATAGATGGTT CTCGTGACTT AGAGGATGTT GTTGTGGATA TCATTGAGAG 24240

ATATACAGAA GAGGTAGCGG CTGATCACTA TGCTAGTCGT GAATTATTGT TTCACTTTAT 24300

TGTGACCAAT ATTAGTTTTC ATGTTAAAGA GGTTCCAGAT TATATAGATG TAACTGACAA 24360

AACTGCAGTT CGTAGCTTTA TGAAGCAGGT GATTGATAAA GAACTTTCTG AAAAGAAAGA 24420

ATTACTTAAT CAACATGACT TATATGAACA GTTTTTACGA CTTTCACTGC TTAAAGCCAT 24480

TGATGACAAC TGGGTAGAGC AGGTAGACTA TCTACAACAG CTATCCATGG CTATCGGTGG 24540

TCAATCTGCT AGTCAGAAAA ATCCAATCGT AGAGTACTAT CAAGAAGCCT ACGCGGGCTT 24600

TGAAGCTATG AAAGAACAGA TTCATGCGGA TATGGTGCGT AATCTCCTGA TGGGGCTGGT 24660

TGAGGTCACT CCAAAAGGTG AAATCGTGAC TCATTTTCCA TAAAAGGAGA AAATATGACA 24720

ATTTACAATA TAAATTTAGG AATTGGTTGG GCTAGTAGCG GTGTTGAATA CGCTCAAGCC 24780

TATCGTGCTG GTGTTTTTCG GAAATTAAAT CTGTCCTCTA AGTTTATCTT TACAGATATG 24840

ATTTTAGCCG ATAATATTCA GCACTTAACA GCCAATATTG GTTTTGATGA TAATCAGGTT 24900

ATCTGGCTTT ATAATCATTT CACAGATATC AAAATTGCAC CTACTAGCGT GACAGTGGAT 24960

GATGTCTTGG CTTACTTTGG TGGTGAAGAA AGTCACAGAG AAAAAAATGG CAAGGTTTTA 25020

CGTGTATTCT TTTTTGACCA AGATAAGTTT GTAACCTGTT ATTTGGTTGA TGAGAACAAG 25080

GACTTGGTTC AACATGCCGA GTATGTTTTT AAGGGAAACC TGATTCGGAA GGATTACTTT 25140

TCTTATACGC GTTATTGTAG CGAGTATTTT GCTCCCAAGG ACAATGTTGC AGTCTTATAC 25200

CAACGAACTT TTTATAATGA AGACGGGACT CCAGTCTATG ATATCTTGAT GAATCAAGGG 25260

AAGGAAGAAG TTTATCATTT CAAGGATAAG ATTTTCTATG GAAAGCAAGC TTTTGTGCGT 25320

GCCTTTATGA AATCTTTGAA TTTGAATAAG TCTGATTTGG TCATTCTCGA TAGGGAGACA 25380

GGTATTGGAC AGGTTGTGTT TGAGGAAGCA CAGACAGCAC ATCTAGCGGT AGTTGTTCAT 25440

GCGGAGCATT ATAGTGAAAA TGCTACAAAT GAGGACTATA TCCTTTGGAA TAACTATTAT 25500

GACTATCAGT TTACCAATGC AGATAAGGTT GACTTCTTTA TCGTGTCTAC TGATAGACAA 25560

AATGAAGTTC TACAAGAGCA ATTTGCCAAA TATACTCAGC ATCAGCCAAA GATTGTTACC 25620

ATTCCTGTAG GCAGTATTGA TTCCTTGACA GATTCAAGTC AAGGGCGCAA ACCATTTTCA 25680

TTGATTACGG CTTCACGTCT TGCCAAAGAA AAGCACATTG ATTGGCTTGT GAAAGCTGTG 25740

ATTGAAGCTC ATAAGGAGTT ACCGGAACTA ACCTTTGATA TCTATGGTAG TGGTGGAGAA 25800

GATTCTCTGC TTAGAGAAAT TATTGCAAAT CATCAGGCAG AGGACTATAT CCAACTCAAG 25860

GGGCATGCGG AACTTTCGCA GATTTATAGC CAGTATGAGG TCTACTTAAC GGCTTCTACC 25920

AGCGAAGGAT TTGGTCTGAC CTTGATGGAA GCTATTGGTT CAGGTCTACC TCTAATTGGT 25980

TTTGATGTGC CTTATGGTAA TCAGACCTTT ATAGAGGATG GGCAAAATGG TTATTTGATT 26040

CCAAGTTCAT CTGACCATGT AGAAGACCAA ATCAAGCAAG CTTATGCCGC TAAGATTTGT 26100

CAATTGTATC AAGAAAATCG TTTGGAAGCT ATGCGTGCCT ATTCTTACCA AATTGCAGAA 26160

GGCTTCTTGA CCAAAGAAAT TTTAGAAAAG TGGAAGAAAA CAGTAGAGGA GGTGCTCCAT 26220

GATTGAACTT TATGATAGTT ACAGTCAAGA AAGTCGAGAT TTACATGAAA GTCTAGGCGC 26280

TACTGGTCTT TCTCAACTTG GAGTGGTCAT CGATGCAGAT GGTTTTCTGC CTGATGGTCT 26340

GCTTTCTCCT TTTACCTATT ATCTAGGTTA CGAGGATGGA AAACCTCTCT ATTTTAATCA 26400

AGTTCCCGTT TCAGATTTTT GGGAAATTTT AGGAGATAAT CAGTCTGCTT GTATTGAAGA 26460

TGTGACGCAG GAGAGGGCTG TCATTCATTA TGCTGATGGA ATGCAGGCTC GCTTGGTTAA 26520

ACAGGTAGAC TGGAAAGACC TAGAAGGTCG AGTACGTCAG GTTGACCACT ACAATCGCTT 26580

CGGAGCTTGT TTTGCTACAA CGACTTATAG CGCAGATAGC GAGCCGATTA TGACAGTTTA 26640

CCAAGATGTC AATGGTCAAC AAGTTTTACT GGAAAACCAT GTGACGGGTG ATATCTTATT 26700

GACTTTGCCA GGTCAGTCCA TGCGTTACTT TGCAAATAAA GTTGAATTTA TCACCTTCTT 26760

TTTGCAAGAT TTGGAAATAG ATACCAGTCA GCTTATCTTT AATACTCTAG CGACTCCTTT 26820

CTTGGTTTCC TTCCATCATC CAGATAAATC TGGCTCGGAT GTCTTGGTAT GGCAGGAACC 26880

TCTCTATGAT GCCATTCCAG GTAATATGCA GTTGATTTTG GAAAGTGATA ATGTGCGTAC 26940

TAAGAAGATC ATCATTCCAA ATAAGGCGAC TTATGAGCGC GCTTTAGAGT TAACTGACGA 27000

GAAATACCAT GATCAGTTTG TGCACTTGGG TTATCATTAC CAGTTCAAAC GTGATAATTT 27060

CCTAAGACGA GATGCCTTAA TCTTGACCAA TTCAGATCAG ATTGAGCAAG TAGAAGCAAT 27120

CGCAGGAGCC TTGCCTGATG TCACTTTCCG TATTGCAGCG GTGACAGAGA TGTCTTCTAA 27180

GCTCTTAGAC ATGCTTTGCT ATCCTAATGT GGCCCTTTAC CAGAACGCTA GTCCACAGAA 27240

GATTCAGGAG CTGTATCAAC TGTCGGATAT TTACTTGGAT ATAAACCACA GTAATGAGTT 27300

GCTACAGGCA GTGCGTCAGG CCTTTGAGCA CAATCTCTTG ATTCTTGGCT TTAATCAGAC 27360

GGTGCACAAT AGACTTTATA TCGCTCCAGA CCATCTATTT GAAAGTAGTG AAGTTGCTGC 27420

TTTGGTTGAG ACCATTAAAT TGGCCCTTTC AGATGTTGAT CAAATGCGTC AGGCACTTGG 27480

CAAACAAGGC CAACATGCAA ATTATGTTGA CTTGGTGAGA TATCAGGAAA CCATGCAAAC 27540

TGTTTTAGGA GGCTAACATG TCAGAGGAAG ATTTATTTTA CAAAGACGTT GAAGGCCGCA 27600

TGGAAGAGTT GAAACAAAAA CCCATCAAGA AGGAAAAAGA AACCCGAGGG GAAAAGATTA 27660

GTAAGACTTT TTCACTTTTA CTGGGTTTGA TGATTCTGAT TGGTTTGCTC TTTACTTTGC 27720

TGGGAATTTT GAGGTAGATC TATGATTGAA ATACTAATTG TTTTAGCTAT TATCCTATCT 27780

CTTGCTTTGA TTGTATTGGT AACTATACAA CCCCGTCAAA ATCAACTATT TTCCATGGAT 27840

GCCACTAGTA ATATTGGTAA ACCAAGCTAC TGGCAGAGCA ACACCTTGGT CAAGGTGCTC 27900

ACTTTATTGG TGAGTTTGGC TTTATTTATT CTACTATTAA CCTTTATGGT GATTACTTAT 27960

AAATAAAAGA AAACTTCAGA TATTCACCTT TTGTGGATTG GTCTGAAGTT TTCTTTTTTA 28020

TACTCAATGA AAATCAAAGA GCAAACTAGG AAGCTAGCCG CAGGCTgCTC AAAACACCGT 28080

TTTGAGGTTG TAGATATAAC TGACGAAGTC AGCTCAAAAC ACCGTTTTGA GGTTGTAGAT 28140

ATAACTGACG AAGTCAGCTC AAAACACCGT TTTGAGGTTG TGGATAGAAC TGACGAAGTC 28200

AGCTCAAAAC ACCGTTTTGA GGTTGTGGAT AGAACTGACG AAGTCAGCTC AAAACACCGT 28260

TTTGAGGTTG TGGATAGAAC TGACGAAGTC AGCTCAAAAC ACCGTTTTGA GGTTGTGGAT 28320

AGAACTGACG AAGctCAGTA ACATATATAC AGCAAGGCGA CGCTGACGTG GTTTGAAGAG 28380

TATTACTGTC TATATTTTTG GTAAAAATCA ACTTTTACTT GGATGAAGGT TTTGGCTTCA 28440

CGTAGGAGTT GAAGAAGGGT GGCGCGGGTT TCAAATTCTT CTCTTGTCTT GGGCAGACTG 28500

CGGTTCCGGA AGACTTCCAG ATAACGTTCA ATTTCATCTA GCAAATCAGA AGCAGGATTG 28560

GTCTGGCTCA GTTGACCTGC AATTTTTGAA AAGAGTTGCG CTAAGATCAG GCTTTCACTG 28620

GCGGCAAGGT GACAAGTGTT AATCTGTTGG GCCATGTTTC TCAGGATACG ACTTTGTCGC 28680

TGTCTCATCT CAAAGTAGTG GATATGGTAG TCTGTCTGGT GAAAGAGGTG GTCAGAGTGA 28740

TCCAAATAGA CCAGTCTGAG GGCTTCTTTC AAAAGCGTGT CTAATTCTGC TACCAGCTGT 28800

GCTCGGTTGC GTCCGTCTCC TCTGGATAAA TAGTATTTGA AGCGCTGGAG GATATCTTTT 28860

AACTTTTCTT CCACCAGCGT GTGGTAGTGC TGGATTTCCT CTTCTCGTGA AGGCATATAG 28920

AGATTAACAA GCAAGGCAAA TCCTGTACCA ATAGCAAAGA GAAGGAATTC ATTGACTAGA 28980

AGGTCTGGAG AGGTTGACTC TTGAACCAAG AGATGGCTAA CCAAAACAGT GCTTGGTGTG 29040

ATGCCAATTT CCCAGCCCAT CTTGTAGGCT AAAGGAACGT AGAAGGCCAG ATAGAGGCCG 29100

AGACTCCAGA TATGAAATCC GCTCAAGTGA AAAGCTAGAA CACCGATAGC CAGAGCTAGA 29160

AGCATAGAAA AAAGACGATT GCGAGCCAGT TTTAAAGTAC TTCTACGCGT ATCAGATAGG 29220

CTCAAGAGAG CGATAATTCC AGCCGAAACT GCTGACGAAA GATTGAGAAA ATAAGCAAGC 29280

AGGCAGGCAA GACAGGTAGC TAAGATGAGC TTGGTCGTAC GTTGGCTAAT AGACATAAGA 29340

ATTTCCTAAT AAGTTAGAAT AAAAGCGTAA AAGACAAGAC ATGAGCAGGC TTGCCTTGAT 29400

GAGTTATTTT TTACGGGTTG CTGCGTATTC GGCAACGGCG GTAAAGAGGA CATCTGTAGA 29460

AGAGTTAAGG GCTGTTTCAC ATGAGTCTTG GATGACACCA ATCACAAAAC CAACCCCAAC 29520

AATTTGTATG GCAATATCGT TAGAAATACC GAAAAGGCTA CAAGCAACTG GGATAAGAAG 29580

GAGGGAACCT CCGGCAATAC CTGAAGCATC ACAGGATGAG ATAGCTGCTA CCACACTGAG 29640

GACAAAGGCT GTGGCAAAGT CAACAGGAAT TCCAAGAGTG TTAACTGCAG CAAGGGTCAA 29700

AAGGTTAATG GTAATCGCTA CTCCAGCCAT ATTGATAGTA GAACCGAGTG GGATAGAAAC 29760

AGAATAGGTA TCTGGGTTGA GTCCAAGGTC ATGGCAGAGT TTCATGTTGA CAGGAATGTT 29820

AGTCGCAGAA CTACGAGTGA AAAAGGCTGT CACACCGCTG ACACGGAGGC AGTTCCAAAC 29880

TAGAGGGTAA GGATTGCGTC TCATAAAGAA GAAGGCAATC AAAGGGTTGA CCACAGGGGC 29940

AACAAAAAGC ATAGTCGTTA CTAATAGAAC CAATAAAATA CCGTAGTTGG CAAGGCTTCC 30000

GACTCCCTTG TCAGAAATGG TTTTAAAAAC AAGACCAAGG ATTCCAAATG GAGCCAGATT 30060

GATGATCCAT TCGACAATTT TAGAAGTCAC GTCAGCGATA GTTTTTAGCA ATTCTTGACT 30120

ATTTTTACTG GCTTCTCTCA TAGCGATTCC AAAAATGACT GCCCAAGATA AGATTCTAAT 30180

ATAGTTAGCA GTAAGCAGGG CGTTGACTGG GTTGTCAACC AGTTTGAGCA AGAGGTTGCT 30240

GAGAACCTGC CCAATCCCAT CTGGTGGTGC AATTTCAGTA TTGGCACTAT TTGGGGTAAT 30300

TTCAATAGGG ACGATGAAAT TTGCTAGTAC AGCTACAAGA GCAGCGGCGA AAGTCCCTAT 30360

CATAGGATAT ACAAGAAAAC AACAGTTTTC ATATTGCTAT CTTGTCCCTT TTGATGTTGG 30420

GAAAGGGCAT TGGCAACGAG AGCAAAGACT AGGATAGGAG CAACAGCTTT TAGACCTCCA 30480

ACGAATAAAT CCTCGAGTAG CCCAATCCCT GAGAGATTAG GAAGGGTCAG TCCTAGGATT 30540

CCCCACAAAG CATACCAATC AAGATACGCT TGACAAGGCT TGCCTTATTC CAAGCATGAA 30600

TGATTCTTTT CATAATAATC TCCTTTTTGT GTAGTGATTA TGATTATAGT ATAAATGATA 30660

GACAAAATCA AGAATTTTCT GTCTATTTTT TGAATATTTA TGGAGAATGA GACTGATGAA 30720

AATATGGTAT AATGAAATAA AGGAGTTTTA TATGCAAAAA TTTATTCAGG CTTATATTGA 30780

AAAGCTAGAT GTGACAACCA TTATCGAGAA TATTCTAACC AAGGTCATTT CTCTTTTACT 30840

GCTTTTAATT GTATTTTATA TTGCTAAAAA AATGCTTCAT ACCATGGTGC AGAGAATTGT 30900

CAAACCTTCT CTAAAAATGT CTCGTCATGA TGTTGGACGC CAAAAAACCA TCTCACGTTT 30960

ACTAGAAAAT GTGTTTAATT ATACGCTATA TTTCTTTTTA CTCTACTGCA TTTTGTCGAT 31020

TTTAGGTTTG CCAGTTTCTA GTTTGCTGGC TGGAGCTGGT ATTGCTGGGG TAGCGATTGG 31080

TATGGGAGCC CAAGGCTTTC TGTCTGATGT CATCAATGGC TTTTTCATCC TCTTTGAACG 31140

TCAACTGGAT GTGGGAGATG AGGTCGTTCT GACAAATGGA CCGATTACTG TATCGGGTAA 31200

GGTTGTCAGT GTGGGAATTC GTACGACACA GCTTCGTAGC GAGGAGCAAG CCCTTCACTT 31260

TGTCCCTAAC CGAAATATCA CAGTTGTTAG CAATTTCTCA CGCACAGACT AGACCTGTTA 31320

TTTTAAGTAA TTTGTGGTAC AATAGAGGGA GTTTAATAAG GAGAAAAGAT GGTTTTAGAA 31380

AAGCAGTTGG GCAATGGTTG TACCTGGATA GACCTAGACC TAGGAAAGTT GAATAAACTA 31440

GAAGACCTTT CTGAAATTTA CGGTTTGGAC AAGGAAACCA TTGAATACGC ACTGGATAGA 31500

AACGAGCGCG CCCACATGGA CTACCACCGT GAAAGTGAGA CGGTTACCTT TATCTATAAT 31560

GTCTTAGACG TAAAAAAGGA CAAGGCCTAC TATGAGACTT TTCCCATGAC CTTTATTGTC 31620

GAGCATCGTC GCCTGATTAC CATTAGTAAT ACCAAGAACG CCTATGTCAT TGAACAGATG 31680

ACTCGTTATC TGGAGAACCA TGACACGCTT TCGATTTATA AGTTTCTCTT TGCCAGTCTG 31740

GAAATCATCA GCAATGCCTA CTATCCTGTC ATTGAGCAGA TGGACAAGAG TAGGGATGAG 31800

GTCAATGACC TCTTGCGCCA GCGAACTACC AAGAAAAACC TCTTTGTCCT GTCTGATTTG 31860

GAGACTGGTA TGGTTTATCT GACGGCAGCT GCCAAACAAA ATCGGATTTT GTTAGAGCAT 31920

ATTCAAGGTC ATGCCTTGTA TCGTAGTTTT GATGAGATTG AGAGAGAACA GTTTGATGAT 31980

GCCATGATTG AGGCTCATCA GCTGGTATCC ATGACAGACC TAATCTCTCA GATTTTACAG 32040

CAGCTTTCAG CCTCTTACAA CAATATTCTA AACAATAATC TGAATGACAA TTTGACAACC 32100

TTGACTATCA TTTCAGTCTT GCTAGCTGTT TTGGCAGTCG TGACAGGCTT TTTCGGAATG 32160

AATGTTCCCT TACCTTTAAC AGATGAGCCC CATGCTTGGC TCTATATCAG TTTGGCTAGT 32220

GCAGGTTTGT GGATTGTTTT ATCCTTGTTA CTAAGGAAAA TTGCGAAAAA AAGTTAAGAA 32280

AAGGAGCCAG AATGGCGATT GAAAATTATA TACCAGATTT TGCTGTGGAA GCAGTCTATG 32340

ATCTGACAGT CCCAAGCCTG CAGGCGCAGG GAATAAAGGC TGTTTTGGTC GATTTGGATA 32400

ATACCCTCAT TGCTTGGAAC AACCCTGATG GAACGCCAGA GATGAAGCAA TGGCTACATG 32460

ACCTTCGGGA CGCGGGTATT GGCATTATCG TAGTGTCAAA TAACACCAAA AAACGCGTTC 32520

AACGAGCAGT TGAGAAATTT GGGATTGATT ACGTTTACTG GGCCTTGAAG CCCTTCACAT 32580

TTGGTATTGA CCGTGCTATG AAGGAATTCC ACTATGACAA AAAGGAAGTG GTCATGGTTG 32640

GTGACCAACT CATGACAGAT ATACGAGCAG CCCACCGTGC AGGGATTCGG TCAATTTTAG 32700

TCAAACCCTT GGTCCAACAT GACTCAATCA AAACGCAGAT TAACCGAACT CGTGAGCGTC 32760

GTGTTATG 32768

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14872 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

CCAGTCACAA AGAAATTGAG CGCGTTCAGc TGAGGATGCA CTATGATGCA AGCTACATTT 60

CATTTGATGG GATATTAAGA AAGGAGATTT TCATGACACT TTTAGATGTA AAACACGTTC 120

AAAAAATTTA TAAAACACGT TTTCAGGGCA ACCAAGTAGA AGCCCTCAAG GATATTCACT 180

TTACCGTAGA AAAGGGTGAC TACGTTGCCA TCATGGGTGA GTCTGGTTCT GGTAAATCAA 240

CTCTTCTCAA TATTCTAGCT ATGTTGGATA AACCAAGTCG TGGTCAGGTT TACTTGAATG 300

GAACTGACAC CGCAACTATT AAAAATTCAC AGGCTTCTAG TTTCCGGCGT GAAAAGCTAG 360

GATTTGTCTT CCAAGACTTT AACTTGCTAG ATACTCTGTC TGTTAAGGAC AATATCTTGC 420

TTCCGCTTGT CTTGTCAAGA AGACCTATAA CGGAGATGAT GAAGAAATTG GTGGTGACAG 480

CTGAGAATCT GGGTATTAAC CAATTGCAAG AGAAGTACCC TTACGAGATT TCTGGTGGTC 540

AGAAACAGCG TGTAGCAGTA GCCCGCGCCA TCATCACAGA ACCTGAAATT CTCCTTGCGG 600

ACGAGCCAAC AGGAGCCCTT GATTCCAAGT CATCTGCAGC CTTACTTGAT GTCTTTAATG 660

AAATCAATGA GCGTGGGCAA ACCATCCTCA TGGTAACCCA CTCAACAGCA GCTGCTAGCA 720

GGGCCAAGCG TGTTCTCTTT ATCAAAGACG GCATTCTTTA CAACCAAATC TACCGTGGAG 780

AGAAGACAGA GCGTCAGATG TTCCAAGAAA TCTCTGATAC CTTGACTGTC ATGGCAAGCG 840

AGGTGAATTA GTATGTTTCG ATTAACCAAT AAGTTAGCGG TATCGAACTT GATTAAAAAC 900

CGCAAACTCT ACTATCCCTT TGCACTGGCT GTTCTCTTGG CAGTCACCAT CACCTATCTC 960

TTTTACTCCC TAACCTTCAA TCCAAAGATT GCGGAAATCC GTGGAGGAAC CACCATTCAA 1020

GCAACACTTG GATTTGGTAT GTTTGTCGTT ACCCTTGCGT CAcCATTATC GTCCTCTATG 1080

CCAATAGTTT TGTCATGAAA AACCGTTCCA AGGAACTGGG TATATATGGC ATGTTAGGCT 1140

TGGAGAAGCG CCATCTAATC AGTATGACCT TTAAGGAGTT AGTGGTATTT GGGATTCTAA 1200

CTGTTGGAGC GGGTATCGGT ATTGGAGCCT TGTTTGACAA GTTAATTTTC GCTTTCCTGC 1260

TCAAACTAAT GAAACTGAAG GTTGAGCTGG TTGCTACCTT CCAAATGAAT GTTGTCATTG 1320

CAGTACTTGT TGTCTTTGGA TTGATTTTCC TAGGCCTCAT GTTCCTGAAT GCTCTTCGAA 1380

TCGCCCGTAT GAATGCCCTC CAGCTCTCGC GTGAGAAAGC AAGCGGAGAG AAAAGAGGTC 1440

GCTTCCTACC TCTCCAAACG ATTCTTGGTT CCATAAGTTT AGGGATTGGC TATTATCTTG 1500

CCCTTACGGT AACCGATCCT CTTACAGCCC TAACAACTTT CTTCCTAGCT GTTTTGCTGG 1560

TTATCTTTGG TACTTATCTA TTGTTTAATG CAGGGATTAC AGTCTTCCTA CAAATCTTAA 1620

AGAAAAACAA GAAATACTAT TACCAACCTA ATAACCTCAT ATCTGTTTCC AACTTGATTT 1680

TCCGTATGAA GAAAAATGCG GTTGGACTAG CAACCATCGC TATTTTGTCA ACAATGGTTT 1740

TGGTAACCAT GTCAGCAGCG ACAAGCATTT TCAATTCCGC AGAAAGCTTT AAAAAAGTTC 1800

TAAATCCTCA TGATTTTGGG GTTTCAGGGC AAAATGTTGA AAAAGAAGAT TTGGACAAAC 1860

TCTTGAGCCA GTTTGCAAGT GACAAAGGTT ATAGTGTCAA AGAGAAAGAA GTACTTCGTT 1920

ACAGTAACTT TGGTATTGCA AATCAAGAAG GAACCAAGTT AACTATTTTT GAAAAAGGAC 1980

AAAACCGTGT CCAACCCACA ACAGTTTTCA TGGTATTTGA CCAAAAAGAT TATGAAAATA 2040

TGACTGGTCA AAAACTGTCT CTATCAGGAA ATGAGGTCGG TCTCTTTGCC AAAAATGACG 2100

GACTGAAAGG ACAGAAAGCT CTAACTCTAA ATGATCATCA ATTTTCTGTC AAAGAAGAAT 2160

TTAATAAAGA TTTCATTGTG AACCATGTTC CAAATAAGTT TAATATCTTG ACTACTGATT 2220

ACAATTACCT TGTTGTTCCT GATTTACAAG CCTTTTTGGA TCAATTCCCA GATTCGGCTA 2280

TCTATAATCA GTTTTACGGT GGTATGAATG TAAATGTCAG TGAAGAAGAA CAACTCAAGG 2340

TCGCTGAGGA GTATGAAAAC TACCTCAATC AATTTAATGC TCAATTAGAC ACAGAAGGTA 2400

GCTATGTTTA TGGTAGCAAT CTAGCAGATG CTAGTTCTCA GATGAGTGCC CTCTTTGGTG 2460

GTGTCTTCTT TATCGGTATT TTCCTATCCA TTATCTTTAT GGTCGGAACT GTTCTGGTCA 2520

TCTACTACAA ACAAATTTCT GAAGGCTACG AAGACCGTGA ACGCTTTATT ATCTTGCAGA 2580

AAGTCGGTTT GGACCAAAAG CAAATCAAGC AAACCATCAA CAAACAGGTT TTAACTGTTT 2640

TCTTCCTTCC TTTGCTCTTT GCCTTCATAC ATCTCGCCTT TGCCTACCAT ATGCTTAGCC 2700

TGATTTTAAA AGTGATTGGT GTACTGGATA CGACTATGAT GTTGATTGTG ACCTTGTCTA 2760

TCTGCGCTAT CTTCCTCATC GCCTATGTGC TGATTTTCAT GATTACTTCA AGAAGTTATC 2820

GCAAGATTGT GCAAATGTAA AAAAGATACC TCGACTTCAA AATCGAGGTA TTTCTTGTAT 2880

TCTAAATGCT GAAAAGTTGT CCGAGCAGGA AGGTAACTCC CATGGTCAAG AGACCAATAG 2940

CAAGGTTCCG AATCATAGCT GTTTTGGTTG GGGCTTTTCC AAGTCTAGCA CTTGTGTAAC 3000

CAGTGAGAAG AAGGGCCACA CCGACAATAA GGACGGTAGC AGGGATGCGG TAATCACTTG 3060

GAAAAATGGT CACTGACAGC ATTGGAGGCA AACTTCTAAG GAAAAAGGCA ACGAAGCTAG 3120

AAATGGCAGC GTGCCAAGGA TTGGTAAATT CTTCATACTC AATCCCATAT TTTTCCTCTA 3180

CCAGAGCCTT GAGTGGATTT TTAAGAAAGA TCTTATTGGT CAAGAGTTGG GCAGAAGTTT 3240

TGAATTCTCC ATTTTGGATA TAAGCAGCAT AGAGGGATTT TTTGGCTAGT TCCCTATCTT 3300

GGTCTAGCAA GAGTTTTTCT CGCGAAACGG CAGCTTCCTC GGTATCTTTT GGAGTTGAAA 3360

CGGATACATA TTCTCCACCA GCCATTGAAA AGGCACCAGC TAAGATAGCC GTAAAACCTG 3420

ATAAAAAGAT AATCCAGATA TTGGTCGTGG CACTGGCAAC TCCGATAACC ACACCAGCAA 3480

TGGAAATAAT TCCATCGTTA GCATCAAGAA CACCCGCACG CAGGATATTT AAACGACCTG 3540

CAAAATTTGA ATCAATTTCG TGATTTGTTT CTGACGCTAA ATTTCAAGTT CAAGTTAGCC 3600

ATCAAGAAGT CTTCTCTGGG TGACTTGTAG TCCAAGCATT TTTTAGGATA GTTGTTAATC 3660

CACTTTTCGA TGAATGCGAC TTCTTTGGGA GTCATTTTCT TGGTTCCCTT AGGTAACCAT 3720

CTACGAATGA GCCTGTTGTG ATTCTCATTA GTTCCCCTTT CCCAAGAGGC ATAGGGATGT 3780

GCATAATAAA TGTGCTCCTC AGAAAATACA TTAGACAAGC GATTGAATTC CGTTCCATTA 3840

TCTGCCGTGA TGGAAAGAAT CTTGTGTTGT TTTAAGATGA GTTTTAGAGC CTGATTGACC 3900

ACATCAGCAC TTTTATTTGG AATCAATCGG ATGATCTGAT GTCTACTTTT TCGATCCGTC 3960

AAGACAAGCA AGCAGTAGTT TTTCGCTCTC GTAAGTAGAA CTGTATCAAT CTCATAATGC 4020

CCATTCTCCA AGCGAAAATT GATAGCTTCA AGCCGCTGTT CGATGGATTG ACCAGCAGGT 4080

TTAAAGTTGG TGCTGGCCTG TTTCTTAAGC GCTTTTCCTT TTCTAGGGTA AAGCAGATCC 4140

TGTTTGCTTA ACCCCAATTT TCCATGATGA ATCCAATAGT AAATGGTTGA AATTCCCACG 4200

TTAACCCCTT TAGCCATCAC CATCATTTCA GGCGAAAATT TTTGGTTATG ATAGTGGAGA 4260

ATCTTTTCCT TTAGTTCCTT GGTCAAGCTT GATTTCTTGA CCGAGCGCTT GCGATTGTTT 4320

TCATAAGACT GTTGAGCATA GTCGGCAGAA TAAACCTCTT TGAAGCGCCC TTTTCCAAGA 4380

CATTGTCGGA CTGTCCCACG CTTGATTTCA GTGTGGATAG TTTGAGGAAC TTTTCCAAGC 4440

AGAGAGGCAA TTTCTCTATT TGATTTCCCT TCTTTTTTCC ATCTTTCGAT TAAGCGACGG 4500

CTATCGATTG TCAAATGTTC GCCTTTTGTA GTATAATGGT TTTGCATCTC TGTGCCTTTC 4560

TTGTGTTTGT GGTTGAACAA CAAGTATAAC ACAGAGGTGT TTTCTTATGC CTACAAGAGC 4620

TATCGGCTAG TTGAACCATC TAATTTTTAG GAGGGCTGGG TGGCTAACTT CATTATAGAA 4680

CTTTCATTTA CGAACATATA GTAAAATGAA ACAAGAACAG AACAAATCGA TCAGGACAGT 4740

AAAATCTATT TCTAACAATG TTTTAGAAGC AGAGGTGTAC TATTCTAGTT TCAATCTATT 4800

ATATTTTTGT TTTTTATCAA AAAATACTTT ACAAGTTCTT AAAAACATGA TATAGTAATA 4860

AAGCTTAGAA AATGAGATGA TGTTTTCTAG CAAATATAAA CCCGAGTAAA AAATGCCTAC 4920

GGACAGGCAG GGTTGAATGC CGAAGCGTGG TTGAAAAGCC ACATTATTGA TAGGGTTAAA 4980

AGCCTACTTT TATAAGTTGA TGTTAGGACA CTTGTCCTAA TTCATAAATT TTTAGTGTGG 5040

TGAAAGCACA CGTCATCTTG TGAAACGATC AATAAAGTAC GTAATATTTG CTACTAGAGA 5100

GTTAGGAAAC ATCGGGAACA GACATACTCA ACAGAAACCA AAATAAACAC GTCAGAAGAT 5160

TGCAGAGCAG GTGAAAACCT GCTCTTTTTT CATGAGTCAA CCTTTAGTTC CTTAGTTTTC 5220

ATAAGGTCCT AAAAATATTG AAAGGAGTAT GTTTTGAAAG AGTTAGATCA AAACCAAGCC 5280

CCAATTTATG AGGCCTTGGT GAAGTTACGC AAGAAAAGGA TTGTTCCCTT TGATGTTCCA 5340

GGTCACAAGC GTGGACGGGG AAATCCAGAA CTTGTCGAAC TCTTAGGAGA AAAATGTGTA 5400

GGCATTGATG TCAATTCGAT GAAACCTTTG GATAATTTAG GCCATCCTAT TTCGATTATT 5460

CGTGATGCAG AGGAGCTGGC TGCAGATGCT TTTGGAGCTA GCCATGCCTT TCTAATGATT 5520

GGTGGAACAA CTTCATCGGT GCAGACTATG ATTCTGGCAA CCTGCAAGGC AGGAGATAAG 5580

ATTATTCTGC CACGAAATGT CCATAAATCT GCTATCAATG CGTTGGTTCT ATGTGGTGCC 5640

ATTCCCATCT ATATCGAGAT GAGTGTAGAT CCTAAGATTG GTATCGCTTT AGGTCTTGAA 5700

AATGACCGAG TAGCACAGGC CATAAAGGAC CATCCAGATG CTAAGGCTAT CCTAATCAAC 5760

AATCCTACTT ACTACGGCAT CTGTTCAGAC CTAAAGGGGT TGACAGAAAT GGCTCATGAA 5820

GCTGGCATGA TGGTTTTAGT AGATGAAGCC CACGGAGCGC ATTTGCATTT CACTGATAAA 5880

CTTCCAATTT CTGCTATGGA TGCAGGGGCT GATATGGCAG CAGTTTCCAT GCATAAGTCT 5940

GGTGGGAGTT TGACCCAAAG CTCCATTTTA CTTATCGGGG AGCAGATGAA TTCTGAATAC 6000

GTTCGTCAGA TAATTAACCT GACCCAGTCT ACATCTGCCT CTTACTTGTT GATGGCTAGT 6060

TTGGATATTT CACGTCGCAA CTTGGCCCTT CGTGGTAAAG AGTCGTTTGA GAAAGTCATT 6120

GAGCTATCTG AGTATGCCCG CCGTGAAATC AATGCTATCG GTGGCTACTA TGCCTACTCA 6180

AAAGAGTTAA TAGACGGTGT TTCGGTTTGC GATTTTGACG TAACTAAGCT GTCAGTTTAC 6240

ACTCAGGGTA TTGGCTTAAC AGGTATCGAG GTTTATGACC TCTTGCGAGA CGAATACGAC 6300

ATTCAGATCG AGTTTGGTGA TATCGGCAAT ATCTTGGCCT ATATTTCCAT CGGCGACCGC 6360

ATCCAAGACA TCGAGCGCTT GGTTGGTGCT CTGGCTGATA TTAAGAGACT CTATTCAAGA 6420

GATGGAAAAG ATTTGATAGC AGGAGAATAT ATTCAGCCCG AGTTAGTGCT GTCTCCGCAA 6480

GAAGCCTTCT ATTCAGAAAG AAAAAGTTTA ACTTTGGATG ATTCTGTTGG ACAGGTCTGT 6540

GGAGAATTTG TTATGTGTTA CCCTCCAGGT ATTCCTATCT TGGCTCCTGG TGAACGCATT 6600

ACACGAGAAA TTGTCGACTA TATCCAATTC GCCAAGGAAC GTGGTTGCTC CCTCCAAGGG 6660

ACGGAAGATC CAGAGGTCAA TCATATCAAC GTTATTAAGA GAAAGACAAA CTATAAGAAA 6720

AGTCAATAGT TTTATCTAAA CTATTTCTTA TTTCAATTTG ATGATTTGGC GATGATTTTA 6780

GAGCACGGCA AAAAGCCCTT GAATTAGAAG CGGTCAATCG CTTAATTTCT ATCAGCTTAT 6840

CAAATCCTGC CTCAAGCCTT TTCTGAGGAT TAGGGTAGCG TGTCAAGAGT TGGTAGGTAT 6900

ATTCTGAATG CTTTCCAACG ATTTTATCCA ACTCAGGAAA GATGATATCA AGACAACGAG 6960

TGTATTGTAC TTTCCAATCA GACTGTTTTT TCTTGAGACG ATGAATATGT CTAGCCAGTA 7020

TTTTTAGTTC TACTTGCCGA TTATCGTGTT GAAATTGTTC ACGATTGGGG TCAGAAAGAA 7080

GTTTAAGAGC GATGCCATGA GCGTCTTTCT TATCCGTTTT AGTTTTGCGA AGTGATAATG 7140

ATTTGGCAAA TTTCTTGATG AGCAAAGGAT TGTAGGTGTA AACTTTATAT CCTTGTTCAT 7200

GCAGGAAGTT CAGTAGATTA AAGGCATAAT GTCCGGTATT TTCAAGAGCG ATGAGACAGT 7260

CTTGGTTGAG CTGTCGAAGA GACAGATCTA AGAGTTCAAA ACCAGCTTTA TTATTTGAAA 7320

AAGTGAGTGG TTTAAGAACA GTTTTTCCTG GAACATTCAA GGCTGTAACA TCGTGTTTAT 7380

TTTTAGCGAC ATCAATGCCC ACATAAAGCA TGGGAGTATC TCCAGATATA GTATTTCAAG 7440

TCTACTGGGT TATCCACGAA CTTTTTGCCT TGTTACCTTA GACGAGATAA AACGTCTATG 7500

CGTTATCAAA CTCATTACCA ATTGAAACAA AAAACTGTGG TTAGAGCCTT TCGGAAATCG 7560

TCAAGCGATT GGAGGAAATG AACTAATCCA CAGTGGCTTA TTCCAAGTAT ACCACTTGGG 7620

CTTTGGCAGT AGCTAACTGC GCTAAATATA ATATAAGGAG AAATAGATGG ATTTATGGTT 7680

TTCTGAAGTT CATACTCCAG ATGTCAAATT GTCTCTGAGA ACAGCCAAGC AACTTTACGC 7740

TGGAAAAAGT GAATGGCAGG ATATCGAAGT CTTGGATACG CCAGCTTTTG GGAAAATACT 7800

GATTTTAAAT GGCCATGTCT TGTTCTCAGA TGCGGATGAT TTCGTCTACA ATGAAATGAC 7860

CGTTCACGTT CCCATGGCTG TCCACCCAAA TCCAAAGAAA GTATTGGTTA TTGGGGGTGG 7920

TGACGGCGGT GTTGCCCAAG TATTAACCCT CTATCCTGAA CTGGAGCAAA TTGATATTGT 7980

GGAACCGGAT GAGATGTTGG TCGAGGTCTG TCGTGAGTAT TTCCCAGACT TTGCTGCAGG 8040

GCTAGATGAT CCTCGTGTTA CCATTTACTA CCAAAATGGG CTACGCTTTT TGCGAAACTG 8100

CGAAGATGAT TACGATATTA TCATCAACGA TGCGACAGAT CCATTTGGCC ATACGGAAGG 8160

ACTCTTTACC AAGGAATTCT ACGGCAATAG TTATCGAGCT CTGAAGGAAG ACGGCATCAT 8220

GATTTACCAG CATGGGAGTC CCTTCTTTGA CGAGGATGAG TCGGCCTGCC GAAGCATGCA 8280

CCGCAAGGTC AATCAAGCCT TTCCAATCAG TCGGGTTTAT CAGGCCCATA TTCCAACTAG 8340

CCCAGCTGGC TATTGGTTGT TTGGATTTGC ATCGAAAAAA TACCACCCTG TCAAAGATTT 8400

TGACAAGGAA GGCTGGAAAA AACGCCAGCT TTTCACAGAA TACTACACTG CAAACTTACA 8460

CGTGGGAGCC TTTATGTTGC CCAAGTATGT TGAGGACATT TTAGAAGAAG AGGAAGGAAA 8520

AAAATGAGTC GTTTACTAGT TATTGGTTGT GGGGGCGTTG CCCAAGTTGC TATTTCAAAG 8580

ATTTGTCAAG ATAGCGAAAC ATTTACAGAG ATTATGATTG CTAGCCGTAC CAAGTCAAAA 8640

TGCGATGACT TGAAAGCGAA GCTAGAAGGC AAAACAAGTA CTAAAATTGA AACTGCAGCA 8700

CTTGATGCTG ACAAGGTTGA AGAAGTGATT GCCCTGATTG AAAGCTACAA ACCAGAAGCT 8760

GTTTTGAATG TAGCTCTGCC TTATCAAGAT TTAACCATTA TGGATGCTTG TTTGGCAACA 8820

GGTGTTCACT ATATCGATAC AGCCAACTAC GAAGCAGAAG ACACAGAAGA CCCTGAGTGG 8880

CGTGCTATCT ACGAAAAACG TTGTAAGGAA CTTGGTTTTA CAGCCTACTT TGACTACTCA 8940

TGGCAGTGGG CTTATCAAGA GAAATTCAAA GAAGCAGGCT TGACTGCTCT TCTTGGTTCT 9000

GGTTTTGACC CAGGTGTAAC TAGTGTCTTT TCAGCTTATG CCCTCAAACA CTATTTTGAT 9060

GAAATCCATT ATATCGACAT TTTAGACTGT AATGGCGGTG ACCACGGTTA TCCATTTGCA 9120

ACCAACTTTA ATCCAGAAAT TAATCTCCGT GAGGTTTCTG CGCCAGGTTC TTACTGGGAA 9180

GATGGGAAAT GGGTCGAAGT CGAAGCTATG TCTATCAAGC GTGAGTATGA TTTCCCTCAA 9240

GTTGGACAAA AAGATATGTA TCTCCTTCAC CATGAAGAAA TCGAATCATT GGCCAAGAAC 9300

ATTCCAGGTG TCAAACGCAT TCGTTTCTTT ATGACTTTTG GTCAATCTTA CTTGACGCAC 9360

ATGAAATGTC TTGAAAATGT TGGACTCCTT CGTACGGATA CCATTAACTT TAACGGCCAA 9420

GAAATTGTTC CAATTCAATT TTTGAAAGCC TTGCTTCCAG ATCCTGCCAG TCTTGGGCCA 9480

CGTACAGTCG GAAAAACCAA TATTGGATGT ATCTTTACAG GTGTCAAAGA CGGTGTCAAA 9540

AAGACTATCT ATATCTACAA TGTCTGCGAC CATCAGGAAT GTTACGCAGA GGTTGGTTCG 9600

CAAGCTATTT CTTATACGAC AGGAGTTCCA GCCATGATTG GGACAAAATT AGTCATGAAC 9660

GGAACTTGGA AACAAGCTGG AGTGTATAAC CTTGAGGAGT TAGATCCAGA TCCATTCATG 9720

GAAGCTTTGA ATGAGTATGG TTTGCCATGG GTTGTGGTTG AAAATCCACA AATGGTGGAC 9780

TAATGAAGTT AGAACAAGTA CCAACACCAG CCTATGTTAT TGACTTGGCC AAGTTAGAAG 9840

CTAATTGCCG CATTCTACAA TATGTACAAG AAGAGGCCGG TTGCAAGGTC TTGCTTGCCC 9900

AGAAGGCATA TTCCCTCTAC AAAACTTATC CCTTGATTAG CCAGTATCTA TCAGGTACGA 9960

CAGCTAGTGG ACTCTATGAG GCCAAATTGG CAAGGGAAGA ATTTCCTGGT GAAGTCCATG 10020

TATTTGCGCC TGCTTTCAAG GATGCAGACT TGGAGGAATT GCTAGAGATA ATGGACCATA 10080

TAGTCTTTAA CTCAGAGAGA CAGTTGCGTA AACACGGTCC GCGTTGTCGA GAGGCTGGTG 10140

TCAGTGTTGG TTTGCGCCTC AACCCTCAGT GTTCAACTCA AGGcAGATCA CGCGCTCTAT 10200

GACCCTTGTG CACCAGGTTC TCGCTTTGGA GTTACTATAG ACAAGATTCC GAGTGATTTG 10260

CTAGATTTGG TTGACGGACT TCATTTTCAT ACCCTTTGCG AGCAGGGAGC AGATGATTTA 10320

CAAACAACTT TGAAAGCAGT AGAAGAACAG TTTGGTCCCT ACTTACATGA GGTAAAATGG 10380

CTCAATATGG GTGGTGGTCA TCATATTACA AGAGAAGGTT ACGATGTGGA TTTGCTGATT 10440

TCAGAAATCA AGCGTATCCG AAAAACTTAC AATCTTGAAA TCTATATCGA GCCTGGTGAA 10500

GCCATTGCGC TTAATGCGGG TTATTTAGCA ACTGAGGTAT TAGATATTGT AGAAAACGGT 10560

ATGGAAATCT TGGTTTTAGA CGCCTCTGCG ACCTGCCATA TGCCTGATGT ACTTGAGATG 10620

CCCTATCGTC CACCTTTGAG AAATGGCTTT GAGTCACAGG AAAAAGCCCA TACCTACAGA 10680

CTTTCTTCTA ATACCTGTCT GACGGGCGAT GTGATTGGTG ATTATAGTTT TGAAAATCCA 10740

GTCCAAATCG GAGACAGACT TTATTTTCAA GACATGGCCA TTTATTCTTT TGTCAAAAAT 10800

AATACCTTTA ATGGTATTGG ATTGCCAAGT CTCTATCTCA TGGACGAACA GGGAGACTGT 10860

AGCTTACTCA AAGCTTTTGG CTATCAAGAC TTTAAAGGGA GATTATCATG ATGGACAGTC 10920

CAAAAAAATT AGGCTATCAC ATGCCAGCAG AGTACGAACC CCATCATGGT ACCCTCATGA 10980

TATGGCCGAC TCGACCAGGA TCATGGCCTT TTCAAGGAAA GGCTGCTAAA AGAGCATTTA 11040

600

CTCAGATTAT CGAGACCATA GCAGAAGGGG AAAGAGTCTA TCTTTTGGTG GAGCAGGCCT 11100

ATCTATCTGA AGCCCAATCC TATCTTGGAG ACAAGGTTGT TTATTTAGAC ATTCCCACCA 11160

ATGATGCCTG GGCGCGTGAT ACTGGCCCAA CCATTCTCGT CAATGATAAA GGTAAGAAAT 11220

TAGCCGTGGA TTGGGCCTTC AATGCTTGGG GAGGCACCTA TGATGGTCTT TATCAAGATT 11280

ATGAAGAGGA TGACCAAGTA GCCAGTCGTT TTGCTGAGGC CTTGGAAAGG CCTGTCTATG 11340

ATGCTAAACC TTTTGTACTG GAAGGAGGCG CAATCCATAG CGATGGTCAA GGAACTATTC 11400

TCGTAACTGA AAGTTGCTTG CTTAGTCCTG GTCGCAATCC TAACTTGACT AAAGAGGAGA 11460

TTGAAAACAC ATTATTAGAA AGTCTTGGTG CTGAAAAAGT TATTTGGCTT CCTTATGGTA 11520

TTTATCAGGA TGAAACCAAT GAACACGTCG ATAATGTTGC TGCCTTTGTT GGTCCTGCTG 11580

AGCTTGTTTT GGCTTGGACA GATGACGAAA ATGATCCCCA GTATGCCATG TCAAAAGCAG 11640

ATCTCGAACT CTTAGAACAG GAAACAGATG CAAAAGGTTG TCACTTCACC ATTCATAAAT 11700

TGCCTATCCC TGCAGTTCGA CAAGTTGTGA CAGAAGAAGA TTTGCCAGGC TACATCTATG 11760

AAGAAGGAGA AGAAAAGCGA TACGCAGGTG AACGACTAGC AGCTTCCTAC GTAAACTTTT 11820

ATATCGCCAA CAAGGCTGTC TTGGTTCCAC AGTTTGAGGA TGTAAACGAC CAAGTGGCCT 11880

TAGATATCCT CAGCAAGTGT TTCCCAGACC GTAAAGTTGT CGGAATACCA GCCAGAGATA 11940

TTCTCTTAGG TGGTGGCAAT ATCCACTGTA TCACCCAACA AATTCCAGAA TAGGAGAAAA 12000

AGATGAGAAA TGTAAGAGTT GCAACCATTC AGATGCAATG CGCTAAGGAT GTGGCAACAA 12060

ATATCCAAAC CGCAGAGCGT TTAGTACGTC AGGCTGCTGA GCAAGGAGCC CAAATTATTC 12120

TCTTGCCCGA GTTGTTTGAA CATCCCTATT TCTGTCAGGA ACGTCAGTAT GACTACTACC 12180

AGTATGCCCA ATCTGTAGCG GAAAATACTG CCATTCAGCA TTTTAAGGTG ATTGCTAAGG 12240

AACTACAAGT TGTTTTACCA ATCAGTTTCT ATGAAAAAGA TGGTAATGTC TTGTATAACT 12300

CTATTGCCGT CATTGATGCA GATGGGGAAG TGCTGGGCGT TTATCGAAAG ACCCATATAC 12360

CAGATGACCA TTATTATCAA GAAAAATTCT ATTTCACGCC TGGTAACACT GGTTTCAAGG 12420

TCTGGAATAC TCGCTATGCT AAGATTGGTA TCGGTATCTG TTGGGATCAA TGGTTCCCTG 12480

AAACAGCGCG CTGTCTTGCA TTGAATGGTG CTGAATTGCT CTTTTATCCT ACAGCTATCG 12540

GTTCAGAGCC AATTTTGGAT ACAGATAGTT GTGGTCACTG GCAACGTACT ATGCAAGGGC 12600

ACGCAGCAGC GAATATTGTT CCAGTCATCG CAGCCAATCG TTATGGTTTA GAGGAGGTTA 12660

CTCCTAGTGA GGAAAATGGC GGACAGAGCT CCAGTCTTGA CTTCTACGGT TCCTCCTTTA 12720

TGACGGATGA AACAGGAGCT ATTCTAGAAC GAGCTGAAAG ACAAGAAGAA GCTGTTCTGT 12780

TAGCTACTTA TGACCTAGAC AAGGGAGCAA GTGAACGCCT AAACTGGGGC TTGTTTCGAG 12840

ATAGAAGACC AGAAATGTAT AGACAAATTA CAGATTAGTG TGGGAGAAAT GAGAGATTCA 12900

TTCTGCTAGA CTAACTTCTT ATTAGTAACT ATAAGATACT ATGGCATCTA GTAAATCGAT 12960

TTTTATGATT CGCTATTCTT GTCTATTGAT TAGTCCGTAT TTTAAAATAT TAGCAAAAAA 13020

GCAAATAGCA GTAACTTCTG TCTATTTGCT TTTCTTTTTT ATAGAATATA TTTCTCAATA 13080

GCACGCGCAA CGCCGTCTTC TTCGTTGCTT GAGGTAACGG CATCCGCAAG AGATTTGATA 13140

TAATCGCTGG CATTTCCCAT TGCAATCCCA AGCCCTGCAA ACTGGAGCAT TTCGATATCG 13200

TTATTAGCAT CGCCCATGGC CATAATCTCT GAGGAATCAA TCTTCAAAAT CTCAGCTAGT 13260

CGTGAAAGAG CAGTAGCCTT TGTCGTTCCA AGCGGCATTG CTTCATAAAT GACAGGCTGC 13320

GAACGAACTC CACTGAATCG TTGGCAAAGC TCTTCAGCAA AACGCTGCTC AAAATCGTCT 13380

GTTTGTTCTT TTGTTCCTAA ACACATACCT TGGAACATCC GGAACTTTCC ACTAGTCGCT 13440

TCTTCAAGAG AAATTTCAGT CAGGTCTGAA AATACTAGTT TAGCATCATT TTCAATAACT 13500

TGATTGGGCT TGTCACCGAG AACAAAATAA TGTGACTCGT CAAAAAGTGT CAACTGAACA 13560

TCACTCTTTT CAGCAAGGTC ATAGAGGTAT TCGATGTCAG CTGGACTCAG TTCTTTCCAG 13620

TCAACTAGAC TCCAATCACT GGTCTGGTGA GTTGAACAAC CGTTGTTAAC AATAATATAT 13680

TCGTTCTGGA GGTCAAGCTC CAGTTTTTTG TAGTAGGGGA GGACACCGAA AAGGGGGCGA 13740

CCCGTACAGA GAACCAGTTT GACACCTTTT TCAATGGCTT TGTGAATAGC AGTAATGTGT 13800

GCTTGTGGGA TTTCCTTGGC TTCATTGAGG AGGGTGCCGT CCATATCCAA GGCTAGTAGT 13860

TTAATCATAG GTCTTCCTCT TTATCTTTGC TATTATTATA GCATATTTTG GAGAAGAAAT 13920

TGATAGAAAG CTTGAGACTA ATTGATTTTA TAGTTTAAGA TGTTTTGATG ACAATTCATG 13980

ATTTGAAGAG GATATTTCGC AAAGATATGC TATACTATGT TTGTCAATGT TGCAACTAGA 14040

CAAATTAAAA AACCAACTTA ATATAATAGT TTTTTTGTAA GTAGGTATGA GTAGCAGATT 14100

ACTCAACTAA TCTGAAGAAT AATGGAGGAA ATATATCATG ATTTTAATGA CAAAAAATAT 14160

AAATCTAACA AATGAAGAAT TAGAGCTGAT ACAAGGTGGA GCAGATCCAT ATGGTAAAAA 14220

TCCTAATGGT AGGTACGATT GGGAAATAGA ACCAGTATTA ACTCTGCTGG TTCATGGATT 14280

TTGTCCCAGA GGCACCTATG ATTCAGGATA TATTGGAGGA GGTAATCATC TTTGCAAAGG 14340

AAGTGCTGCG AGATTTTAAG TAAAATTTAT TAGGAATATG AAGAAACAAG GGGAGAAAAC 14400

AGAGGATTTA ATATGAAAAA ACGAGCTATT CAAATTTTAC TAGCATTGTC CTTAATTTTT 14460

TACAAATCAA CTTGGTTTTG GAGGCTTTTC AATTATCTCG CAAAGCCCTA TCTACCAGCA 14520

AGTCGTGAAT TTTTTCAGAT TCTGCTTTTG ATGGAGAGCG GAGTTCTTTT CTTAGCGGTC 14580

ATCTATCTAC TGGTTTTTGC AGGAAAGAAA ATTTTTCATT TCAAGTGGCA GCTGAGGTAC 14640

TTCATCTACC TTTTACTGGG CTACATCATT TCATATATGT CTGACTTCCT CTTTTCGTAT 14700

TTCATATCCC TGTCTTCAAA TCAGATTTCT TTGAATGAAA CGGTAGAAAT GATGGGGAGA 14760

CAGGAGTTCC CTTATGTCTT GCTCATCGTT TGCTTCATCG CCCCTATTGC TGAGGAATTG 14820

ATTTATCGAG GtGTGCTTAT GACAACCTGT TGCAAAAACT CACCTTGGTA CG 14872

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10223 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

CGTGCTATCG GTCTCAAAAC CAATCTGGTC GCTATGGTCA AATCCAGTTG GAAAATCCAT 60

TCTTCTTGGA GCCATCTGCT GGATTGCCAT CATCCTCACC ACTCTTGGTA TGCAGACCCT 120

TATCGGCATT TTCTAATACT CTTCGAAAAT CTCTTCAAAC CACGTCAACG TCGCCTTGCC 180

GTAGGTATAT GTTACTGACT TCGTCAGTTC TATCTGCAAC CTCAAAACGG TGTTTGAGCT 240

GACTTCGTCA GTTCTATCTG CAACCTCAAA ACGGTGTTTT GAGCTGACTT CGTCAGTCGT 300

ATCTACAACC TCAAAACAGT GTTTTGAGCT GACTTCGTCA GTTCTATCTG CAACCTCAAA 360

ACAGTGTTTT GAGCAGCCCG TGGCTAGTTT CCTAGTTTGC TCTTTGATTT TCATTGAGTA 420

TAACACAAAA GGTAGCCCAT CAGCTACCTT TTTCTTATGC TTCCTCAATC AAGCGAGTAT 480

GTTCTCTCTT GATACAGCGA TTCATCACGA TATCATCACA TCCACCATCA CGCAAAATCT 540

CTTTCGCTTC TAAACTTTCA AGTCCTAGCT GTGCCCAAAA AATCTTGGCA TCAGCTTTGA 600

GAAAATCACG CGCCACATCG GGCAGAAATT CACTGCGACG ATAAACATTG ACAATATCTA 660

CAGGAAAAGG AATTTCAGCG AGGCTAGCAT AAGCCTTTTC ACCCAAGATT TCGCCACCTG 720

CCGCCTTGGG ATTGACTGGG ATGATTTTAT AGCCCCGAGC CTGCATTTCC TTTGTTACTC 780

GATTGCTGGT TGTTTCTTCA CGGTCAGACA AACCCACCAC AGCAAGGGTT TTACTCGTTG 840

CGAGATACTG ACGAATCACG CCATCACTTG GATTGATAAA TTCTTGACTC ATAGAAATCC 900

TCCTTTTTCA TCAGTATAGC ACATTTTGAA AAGGTTTGCA GAATTATACT ACAAAAAAGG 960

AGGACTAGCC CCCTTTTTAT TTAGCCTCGT ACCAGGTTGC CCCTTCATTC TCATCTGCGA 1020

TAAGAGGAAC ACTGAGTTGA ATGGCTTCTT CCATGGTTTG TTTCACCAAT TTTTTCATCT 1080

CTACCAATTC AGATTTAGGC ACTTCAAGGA CGATTTCATC GTGCACTTGT AACAGCATCT 1140

TAGTCTGATA ACCACCTGCA ACCAAGGCTT TATCCAGCTG AATCATGGCA ATCTTGAGAA 1200

TATCTGCTGC CGAACCCTGG ATAGGTGAGT TGATAGCAGT TCGCTCCGCA AAACCACGAA 1260

TATTGAAGTT GCGCGAATTG ATATCTGGCA ACTCACGGCG ACGCTTAAAG AGGGTCTCTA 1320

CATAGCCCTT ATCACGCGCC TCCCGCACCA CTTCATCCAT GTAGTTTTTA ATACCTGGAA 1380

AACGTTCAAA GTAGGTATCA ATGTAGGCTT TGGCTTCCTT ACGACTAATT CCCAAATTAT 1440

TAGACAAGCC AAAGTCTGAA ATCCCATAAA CCACTCCAAA GTTAACTGCC TTGGCATTGC 1500

GACGGTCGTT TGCAGTCACA TCATCAGGAC GCTCAATGCC AAAGACCCGC ATGGCTGTCG 1560

AAGTATGGAT ATCTGCCCCC TCTTGGAAGG CCTTAATCAA GTGCTCATCC TTAGAAATAT 1620

GCGCCAAAAC GCGCAATTCA ATCTGTGAAT AGTCAGAGCT GAGTAGCACA CTATCCTCCC 1680

ACTCTGGCAC AAAAGCCTTC CGAATCAAGC GCCCCTGTTC CAATCGGGCA GGAATATTTT 1740

GCAAGTTTGG ATCCACACTA GACAAACGCC CGGTCTGGGT CAAATCCTGC ACATAGCGAG 1800

TATGAATCTT TCCATCAGCC AAAATCCAGT CCTGCAAGCC AATTACATAA GTAGATTGAA 1860

TCTTAGCAAT TTGACGGTAA TCCAGGATTT TCTTAACAAT CGGAGCAATA GGAGCGAGAC 1920

GCTCTAAAAC ATCCACTGCT GTCGAATAAC CTGTCTTGGT TTTCTTAGTG TATTCTAGAG 1980

GAAGTCCCAA TTTCTCAAAG AGAAGCACGC CCAACTGCTT AGGCGAGTTG ACATTAAACT 2040

CCTCACCAGC CAGCTCGTAA ATCTCTTGAG TCAGTTTTTC AATGACAAGC TCATTTTCAG 2100

CCTGCATCTC AAGCAAGGTC TCTTTCTTGA CCATAATCCC AGCAATTTCC ATCTTGGCAA 2160

GGACAAAAGC CAGAGGTTGC TCCATATCAT AAAGAAGCTC TAATTGCCCA TTTTCGCTGA 2220

GTTTTTCAAG TAAAATAGGC TCTGTTTCTA CCAAAACAGC AAGTTTACAA GCTAAGTGTT 2280

CCAAGAATTT CTCACGTTCA GGAATGGCCT TTTTAACACC CTTACCGTAG AAAGTTTCAT 2340

CATCAACCAA GTAAGTCTGA CCATAAAGAC TAGCGATGGT CGCAATTTCA TTGTCCTCCA 2400

CAGTCGAAAG GAGGTATTTA GCCAAACGGA TGTCAAAAGC AGGCGCCTGC AAATCCACAC 2460

CAAAACGTTG CAAAAGAACT TTAACCTTCT TAAAGTCATA AACTCTCAGA GATGTTTTTT 2520

CTAAGAAATC CTTGAAAATC GGGTCTTGCA ACAGCTCAAG CTTGTCTGTG GCATAGAGCT 2580

TATCCCCACA AGACCAGACA AATCCAACCA AATTATCCGT ATGGTAATTC TCACCAAAAA 2640

GCTCAAAGTG GAAGATAGAC TCTTCACTCA GCATATCTTG ACTGATTTGG TCAACAATAG 2700

TAAAATCCAA ACTCTCAGAC ACATCAGCTG ACGACACATT TAAAGCCTGC TTTAGCTGTT 2760

TGAAGCCCAT CTCATCGTAG AATTTCCCAA GATTTTCAAC ATCTGGACCA CTATAGACCA 2820

AGTCCTCTAA ACCAATCGCA ATCGGTGCCT TGGTATCAAT GGTCGCTAGT GTTTTAGACA 2880

AAAAGGCCTG TTCCTTGTCA TTGATGAGAT TTTCCTTCAT CTTAGAAGTC TTCATTCCAT 2940

CAATATTTTC ATAAATCCCC TCAAGCGAAC CATGCTCCAG CAAGAGCTTA ATACCCGTCT 3000

TTTCACCGAC TTTGGTCACC CCAGGGATAT TATCCGACTT ATCACCCATG AGCGCCTTGA 3060

GATCGATAAA CTGAGCTGGT GTGAGGCCCA TTTCTTCCAT GAGGTAATCT GGCGTAAAGG 3120

CCTCAAACTC AGCCACACCT TTCTTGGAAA TTTCAACCAC CGTATGCTCA TCCGTCAGCT 3180

GAATCAAATC CTTGTCCCCA CTGACAATAG TAATATCAAA ACCATCCTGC TCTGCTAGCT 3240

TATCCAGCGT CCCAATGATG TCATCCGCCT CATACTGAGC CAGATCATAG TGACGAATCC 3300

CCATATGATC CAGCAACTCA CGAATGAAAG GAAATTGCTC ACGAAACTCA TCAGGAGTCT 3360

TGGCCCGACC ACCCTTATAG TCCGCATACA TCTCTGTCCG GAAGGTCGTC TTTCCCGCAT 3420

CAAAAGCCAC CAAAATATGA CTCGGCTCAA CCCGCTCCAA TAAATGACTC AACATCAACT 3480

GAAAACCATA AATCGCATTG GTATGCAAAC CAGCCACATT CTTAAAACGG TCCAACTGCT 3540

GATACAGCGC AAAAAACGCC CGAAAAGCTA CAGAAGACCC ATCAATCAAT AATAATTTTT 3600

TCTTATCCAT ACACCCATTA TAAAGGAAAG AATCAAAAAA TACCATTGGG AAGAGCTAGA 3660

GCAAGTATTT TTCAAACTTT TTCCGAATAA ATAGATAGAG CCAGAGAATT TAGTAAACCT 3720

AGATTTAAAA ATGTGCTATA ATATAGTATA TTGAATCTAT AATAGTACAC CTTGACTGCT 3780

AAAATATTTC TATAAATTAA TTTGACTTTC CTGATAGAGT TATTCACATC TTATTTCAAC 3840

TCACTATAGA AGGAGGAATA GGAGGATTCT CAGACATCCG GGCATCAGCC CAACTAATGA 3900

TTTGATTGCT AAGAAAATAT TCAGCAATCC AGAAATCACT TGTCAATTTA TTCGCGATAT 3960

GCTGGACTTG CCAGCAAAAA ATGTGACCAT TTTGGAGGGA AGCGATATTC ACGTATTACT 4020

CTCCATGCCT TACTCGGTGC AGGATTTTTA TACCAGTATA GACGTCTTGG CGGAGTTGGA 4080

TAACGGTACT CAAGTAATTA TTGAGATTCA AGTCCATCAT CAGAATTTTT TCATCAATCA 4140

CTTGTGGGCT TACCTGTGCA GTCAGGTTAA TCAAAATCTT GAAAAAATTC GTCAGCGAGA 4200

AGGTGATACT CACTAGAGCT ACAAACACAT CGCTCCTGTT TACGCCATTG CTATCGTGGA 4260

TAGTAATTAT TTCTCAGATG ACCTGGCTTT TCATAGCTTT AGTATGCGCG AAGACACAAC 4320

AGGTGAGGTA TTGGCGATTA CCAACAATGG ACAGGAAAAC CATCTGGTTA AGATGGCATT 4380

CTTGGAATTA AAAAATACAG AGAAACCAGC AAAGACAAGG TTCGCAAGCC ATGGTTGGAG 4440

TTTTTCGGCA ACAAGCCCTT TACCCAGCAA CCGCAACGAG CCATTACCCA AGCAAATCAA 4500

CTGCTGGACT ACAAGAGCTG GTCCGAGGAG GACAGGAAAA TGTTTAGTCA ACTACATATG 4560

CGAGAAGAAC AAGTCTTGTT AGCACAGGAC TATGCCTTGG AAACTGCTAG GGCTGAAGGC 4620

CTTGAACAAG GACTAGAGCG TGGGAAAGTT GAAGGAAGGG CAGAAAGGAA ACTTTTTGCC 4680

TTCCTAGACA TAGTACGCCA AGGTCTTCTG ACTTCTGAGG TTGCCAGCCA GCAATTAGGT 4740

ATGTCAGTAT CTGAATTTGA GGCACTGTTG TAAAATGGCT CCATAATATC CATAGTGGGT 4800

AAATCCCCTA TGGATATTAT GGAGCCTATT TTGTGTAGAA AAAAAGTCCC ATATGACCTA 4860

TAATGAAAAG CGACAAAACA ACTCATTAGA AAGAATCATA TGGAACAATT ACATTTTATC 4920

ACAAAATTAC TAGACATTAA AGACCCTAAT GTCCAGATTT TAAACATCAT CAATAAGGAT 4980

ACACACAAGG AAATCATCGC CAAACTGGAC TACGACGCCC CATCTTGCCC TGAGTGCGGA 5040

AACCAATTGA AGAAATATGA CTTTCAAAAA CCTTCTAAAA TTCCTTATCT TGAAACGACT 5100

GGTATGCCTA CAAGAATTCT CCTTAGAAAG CGTCGATTCA AGTGCTATCA CTGTTCAAAA 5160

ATGATGGTCG CTGAAACTTC TGATGACGTA CAGTCATATT TCTTCTCTTT TTATTATATC 5220

ACAGTTTTAA ATCTAGCTTT ACTAGATTCA CCGCTACTAT CTATTTATTC GGAAAAAAGA 5280

CGAAAAAACC TGAGAATCAT CTCAGGCTTG GTCATTAAAT TTTTTTCTCA ATATCGAAAA 5340

GTGGAGAAAG TGGTCGTTTT TCATGAATAC GTACGATAGC ATCCCCTAGG AGATGAGCGA 5400

TTGAAATCTG CTCAATCTTA TCAATCAAAC GCTCTTCTGG CAGATAGATG GTATCCAAAA 5460

CAACCAATTT CTTAATAGCT GATTTTTGGA TATTGTCCGT AGCAGGACCA GAAAGAACTG 5520

GGTGCGTACA GCTTGCATAG ACTTCAACAG CACCAGCTTC CGCAAGAGCA TCTGCCGCAT 5580

GACAAATCGT TCCAGCGGTA TCAATCATAT CATCAATCAA GATACAAGTC TTGCCTTCAA 5640

CCTTACCGAT GATATTCATA ACTTCACTAG TATTCATCTT ATCAACGCTA CGACGTTTAT 5700

CAATAATAGC GATAGATGTT TTCAAAAATT CTGCCAACTT ACGAGCACGA GTCACCCCTC 5760

CATGGTCCGG GCTGACAACC ACATAGTCAG AACCAACCAT ACCACGACGC TCAAAATAAT 5820

CTGCAATCAG AGGAGCACCC ATCAAATGAT CCACAGGAAT ATCAAAGAAT CCTTGAATTT 5880

GCGCAGCATG CAAGTCGATG GTCAATAAAC GATCCACTCC AGCTACTTCA AGCATATTTG 5940

CGACAAGTTT TGAAGTGATT GGCTCACGCG CTCTCGCCTT TCTATCCTGA CGTGCATACC 6000

CATAGTAAGG CATGACAACA TTGACAGATT CTGCACTCGC ACGCTTCAAA GCATCTACCA 6060

TAATCAAAAT TTCAAGCAGA TTGTCATTTA CAGGCGAACT AGTTGATTGT AAGATAAAGA 6120

CGTGTTTCCC ACGGATTGAT TCTTCAATGT TGACCTGAAT CTCTCCATCT GAAAATTGGC 6180

GAACACTTGA TTTCCCCAAC TCTATCCCAA TCTCCTGCGC CACACGTTCT GCCAATTCTT 6240

TATTAGAAGA AAGGGCAAAC AGCTTTAAAT CAGAAAAAGA CATGATTTCC TCCGGTATAT 6300

ATGTATAACT TGTGCTTTTC ACAAGATTTT CCATCTACCA TTGTAGCGCT TTTTGCACTA 6360

TTTTTCAATC AAAAATAAAA GAAGGGCACC ATATTTGTAC CCTTGCATCA TTCTTTTGAA 6420

AAATATTCTA GGTCATCAAC TCATTGTGTT TCTCAACAAA GCAATAAGCA TGATAAAAAC 6480

CATAGAGAGC AATAGCCGTA ACCACTGGAA TCGCTAAAGG CAACTCTGTT TCCAACTCCA 6540

CAAAAGGAGA GTTAAACAAG AAGTGAGTTC CCAAGGCTAA ACCTAGAAAA ATAAGGCCCT 6600

GTTTCTTGCC AACCTTCTGT CCTTTATAGG CTCTGTAAAG CAAGTAAACA CCTACTACAG 6660

CTAGACCTGA AAAAGTCCAG TGAGAGGCAA TTCCTGAGAT GATACGCTCT AAAATTCGCG 6720

AAATAGTAAA GTCAAAGCCC TCTGGCAAAT CCGTACGAAT ATAACCAATA TCCTTAATCA 6780

TTTGGAATCC CAAACCGGAA GCAATTCCAA GTAAAAACAA AGATTTTAAT TTTCGCACAG 6840

GAATCAAAGC CAAAACAAAA ACAAGTGACA ATAATTTCAA GGGTTCTTCT ACCAAAGGAG 6900

CCGCAATAGC ACTTTCAAAG GCATTTAAAA ATGGACTATC TGGGAAAAGA ACCCCCAGTA 6960

AATCATGGAT ATAAGTATTA GCAAAACTAG ACAACCAGCC TGAAAGGAAC ATCCCTCCCA 7020

ATAAAGACAG AATCAAAACC TTCTTTGGCA ATTCCCATTT TTCCCAATAC GGAAGAGAAA 7080

ATAAAGAGCC GGAATCATGT AAAAGAGAGC TAGAAAGATA GAAACTCCCA TTAGTCCATA 7140

TTCCGCACCT GACCTCGAAC CGTCCGTATA GTAGATGGTT TCATACTGTA AACCAATACA 7200

TAGCAATAAA ATAAAAATAA ATAAAATATT GCTTTTCTTC ATACACTTTC TTTCTAAATG 7260

AAGTATTTAT AATTCTACGA CTGTCATACT TCCTGTATCA ACATTGTAAA TGGCACCAGA 7320

GATAATGACA TCGTCTGGTA TTAGGGGAGA CTCGATAAGC AGTTGCATAT CCTCGCGTAC 7380

ACTCTCTTCT ATATCTTGGA AGGGCAAGAA GTCCTGGTCT GACACATCGA CACCCAATTC 7440

TTCCTTCAAA TACTCCTGAA AAGGTTCATT TTCAAAGGTC TGAGCACCAC AGTCTGTATG 7500

ATGCAATACC ACAATTTCTC TTGTCCCCAT TTGTTGCTGG GAAATAACTA GAGAACGAAT 7560

CATATCCTCA GTCACTCGAC CACCTGCATT CCGCAAAATA TGAGCATCCC CAAGTGCCAA 7620

ACCTAGAGCT TGCGCAACGT GCAAACGTGA GTCCATACAG GTCACAATGG CTACTCTGGT 7680

TTTAGGTTTA AGTGGCAGAT TTAACTGCCC ATGTAGGGCA ACATAAGCCT GATTGGCTTG 7740

CATAAACTGT TCAAAATACG ACACGATTCC CTCCTTGAAA ATTTGATAGT CAAATATTTC 7800

TCCTATCTTA TCATTTTTAA GAGAATTTGT CACGGATTAT GCAAAGACCT TTTTCAAGAC 7860

TTCCTGAATC GTTGTCACGC CAATGACCTG AATTTCCTTA GGCAGAGTGA TTCCTGTCAA 7920

GGAATTCTTA GGTACATAAA TCTTAGTAAA GCCCAGTTTA GCAGCTTCGT TGATGCGTTG 7980

CTCAATACGA TTCACGCGCC GAATCTCTCC TGTCAAGCCC AGTTCTCCGA CAAAACATTC 8040

CTGAGGATTA GTTGGCTTGT CTTTGTAGCT CGAAGCAATA GCAACTGCAA CAGCCAAGTC 8100

AATCGCAGGT TCATCCAATT TAACACCACC AGCAGATTTG AGATAGGCAT CCTGATTTTG 8160

CAAGAGAAGC CCTGCCCGTT TTTCCAAAAC AGCCATAATC AAGCTAGCAC GGTTAAAATC 8220

AAGTCCTGTC GTAGTACGCT TGGCATTTCC AAACATGGTC GGTGTTACCA AAGCCTGAAC 8280

CTCCGCCAAA ATCGGACGCG TCCCTTCCAT GGTTACAACG ATGGAGGAAC CAGTCGCCCC 8340

ATCCAAACGC TCTTCTAGGA AAACTTGACT CGGATTGAGT ACCTCAACCA AGCCGCCCGA 8400

CTGCATCTCA AAAATCCCAA TCTCATTAGT GGAACCAAAA CGATTTTTGA CCGCTCTCAA 8460

AATACGAAAG GTGTGGTGAC GCTCCCCTTC AAAGTAAAGC ACCGTATCCA CCATATGCTC 8520

CAACATACGA GGCCCAGCCA AGGTTCCTTC TTTGGTCACA TGACCTACGA TAAAGATGGC 8580

AATGTTATTG GTCTTGGCCA ACTGCATGAG TTCAGCGGTC ACTTCACGCA CCTGAGAAAC 8640

AGACCCCTGC ACCCCTGAAA TCTCAGGAGA CATGATGGTC TGGATGGAAT CAATAATGAG 8700

AAAGTCTGGC TGGATACGCT CCACTTCTGC ACGAACACTC TGCATATTGG TCTCTGCATA 8760

GAGATAAAAC TCACTATCAA TATCACCTAA GCGCTCTGCA CGTAGTTTAA TCTGCTGGGC 8820

AGACTCCTCC CCACTGACAT AGAGAACTGT CCCCACTTGG GACAACTGGG TTGAGACTTG 8880

TAGGAGAAGA GTTGATTTCC CAATCCCAGG ATCCCCACCG ATAAGGACGA GACTTCCTGG 8940

TACCACTCCG CCTCCAAGCA CACGGTTGAA TTCCTCCATC TCCGTCTTGG TTCGATTGAC 9000

ATTGATGGAA GTCACCTCAG CTAGTTTCAT GGGCTTGGTT TTCTCACCTG TCAAGGACAC 9060

ACGCGCATTC TTAACTTCGG CAACCTCAAC CTCTTCCACA AAAGAAGACC AAGACCCACA 9120

GTTGGGGCAA CGTCCCAGAT ATTTAGGGGA ATTATACCCA CAATTTTGAC ATACAAATGT 9180

CGCTTTTTTC TTTGCGATGA CAAACCTCTT TCTATATCTC TAACTCACAC TCAATCACTT 9240

GGCAAAAATC AATCTTCTCA TTTGGCACAA ACTGGCGCAT GAGCATTCGA TGAGCAACAA 9300

CTACCACAGT CTGATGTTCT CGATACTTAG ACATACATTC TAGAAACCGA GACTTCATTT 9360

CCGTAGCTGT CTCATATTGA ATAGGACTAT TAGGAAGCAA CTCCCCCTTG TTTTCTAAAA 9420

ACAGTCTTCT AGCTGTTTCA AAGTTTTCTA TTCCTGTTTT ATAGACCTGC CATTCATGTA 9480

ATAAAGGCTC TACTCTTAAA GGAAGACCCG TAGCACAGAC CACATACGAA GCCGTTTCTA 9540

AAGCTCTTGT GACTGCAGAA GATACGATTA TTTCAGCTGA CGAGAGTAAA GGATTTTTGC 9600

TCAATTTCTG GACTTGCTGC CGTCCCATCT CAGACAAGGG TGCCAAATCT ATCCCAAATC 9660

CTATATAAGA ACGCTCCTCT AACTCACGGT AATCTGGCTC CCCATGACGT ACAAAGATAA 9720

TCTTCATTCT AGTGCCCTGT CGATCCAAAT CCACCAGTTC GAACGCCATC AGCTGCATCT 9780

CCATCTGCAA TTAAGAAAGT AGCAAAAACA GCCTGGACAA TACGCTCCCC AACTTCAAGA 9840

ACAACCTCTT GGTCTGTGAT ATTCTTCATC TGCGCAAAAA TATGCCCTTC ATTTCCAGGA 9900

TTTCCATAAT AATCCCCATC AATGACTCCA ACTGAGTTAA TTAAAACCAA GCCCTTCTTA 9960

CGAGGATTTG AAGAACGATC ATAGAGGTAG AGAACCTCAG TCGGCTGCAT ATAAGCCTTA 10020

ACCCCTGTCG GAACCAAGAC AATCTCTCCT GGCGCAACAA CTGTACGCAC AGCAACCTTT 10080

AAGTCGTAAC CAGTCGCATG CGCTGTCTCA CGCTTGGGCA ATAAATTTTC ATCTGTAAAA 10140

CTCGAAACCA ATTCAAAACC ACGAATTTTC ATAATTTTCT CTTTTCTATT ATCATTTATT 10200

CTAGATTATT CTATACTTAT TTA 10223

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16535 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

TGGTTCTGTC CTTATCGGCG CCTTGTCTTG CTTGCCATGG CTACACCAAC TATCTCATCC 60

GACGAAAGTA CACCAACCAC TAACGAACCC AACAACAGAA ATACAACCAC CCTTGCCCAA 120

CCTCTTACTG ATACAGCAGC TGGCTCTGGT AAGAACGAAA GTGATATTTC TTCACCTGGA 180

AATGCAAACG CTTCCCTAGA GAAAACAGAA GAAAAACCTG CTGCAAGCCC AGCCGATCCA 240

GCACCACAAA CTGGACAAGA TCGTTCAAGT GAGCCAACTA CTTCTACTAG TCCAGTAACA 300

ACTGAAACTA AGGCAGAAGA GCCCATCGAA GATAACTACT TCCGTATCCA TGTCAAAAAA 360

CTTCCTGAAG AAAACAAGGA TGCTCAAGGA CTATGGACTT GGGACGATGT TGAAAAACCA 420

TCTGAAAACT GGCCAAACGG AGCTTTGTCC TTCAAGGATG CCAAGAAAGA TGACTACGGC 480

TATTACCTAG ATGTCAAATT AAAGGGAGAA CAAGCCAAGA AAATTAGCTT CCTCATCAAC 540

AATACAGCTG GAAAAAATCT AACCGGCGAT AAATCTGTAG AAAAACTAGT TCCAAAAATG 600

AACGAAGCTT GGTTAGACCA AGATTACAAG GTTTTCTCTT ACGAGCCACA GCCTGCAGGA 660

ACTGTTCGCG TCAACTACTA CCGCACAGAT GGCAACTATG ACAAGAAATC TCTCTGGTAC 720

TGGGGAGATG TGAAAAATCC AAGTAGCGCT CAATGGCCTG ACGGAACAGA CTTTACGGCT 780

ACAGGCAAAT ATGGCCGCTA TATCGACATT CCTCTTAATG AAGCCGCAAG AGAATTTGGA 840

TTTTTATTAC TAGATGAGAG CAAACAAGGA GACGACGTGA AAATCCGTAA AGAAAATTAT 900

AAGTTCACAG ATTTGAAAAA TCATAGCCAA ATTTTCCTAA AAGACGATGA TGAATCGATT 960

TACACAAATC CATACTATGT CCATGATATC CGTATGACAG GAGCCCAACA CGTAGGCACT 1020

TCTAGCATTG AAAGTAGCTT TTCAACACTT GTCGGTGCTA AAAAAGAAGA TATCCTCAAA 1080

CACTCCAACA TCACTAATCA CCTAGGAAAC AAGGTAACTA TTACCGATGT TGCAATCGAT 1140

GAAGCTGGTA AGAAAGTGAC CTACAGCGGA GATTTCTCTG ACACAAAACA TCCTTATACT 1200

GTTAGCTACA ATTCCGACCA ATTCACTACC AAAACAAGCT GGCGCCTGAA AGATGAGACA 1260

TACAGCTATG ATGGCAAACT GGGAGCTGAC CTAAAAGAAG AAGGAAAACA AGTTGATTTG 1320

ACCCTTTGGT CACCAAGTGC TGATAAGGTT TCTGTTGTTG TCTACGACAA GAATGACCCT 1380

GACAAAGTAG TTGGAACTGT CGCTCTTGAA AAAGGGGAAA GAGGAACTTG GAAACAAACT 1440

CTAGACAGCA CAAACAAACT CGGAATCACA GATTTCACTG GCTACTATTA TCAATACCAA 1500

ATCGAGCGTC AAGGTAAAAC TGTTCTTGCA CTCGATCCTT ACGCTAAATC TCTTGCTGCT 1560

TGGAATAGCG ACGATTCCAA GATTGACGAT GCCCATAAAG TGGCTAAAGC CGCCTTTGTA 1620

GATCCAGCTA AACTCGGACC TCAAGACTTG ACTTATGGTA AGATTCACAA TTTCAAGACT 1680

CGTGAAGACG CCGTTATCTA CGAAGCTCAT GTGCGTGATT TCACTTCAGA TCCTGCCATT 1740

GCAAAAGACT TGACCAAACC ATTTGGGACT TTTGAAGCCT TCATTGAAAA ACTAGACTAT 1800

CTCAAAGACT TGGGTGTAAC CCATATCCAG CTCCTTCCAG TCTTGTCTTA CTACTTTGTC 1860

AATGAATTGA AAAACCATGA ACGCTTGTCT GACTACGCTT CAAGCAACAG CAACTACAAC 1920

TGGGGATATG ACCCTCAAAA CTACTTCTCC TTGACTGGTA TGTACTCAAG CGATCCTAAG 1980

AATCCAGAAA AACGAATCGC AGAATTTAAA AACCTCATCA ACGAAATCCA CAAACGTGGT 2040

ATGGGAGCTA TCCTAGATGT CGTTTATAAC CACACAGCCA AAGTCGATCT CTTTGAAGAT 2100

TTGGAACCAA ACTACTACCA CTTTATGGAT GCCGATGGCA CACCTCGAAC TAGCTTTGGT 2160

GGTGGACGCT TGGGGACAAC CCACCATATG ACCAAACGGC TCCTAATTGA CTCTATCAAA 2220

TACCTAGTTG ATACCTACAA AGTGGATGGC TTCCGTTTCG ATATGATGGG AGACCATGAC 2280

GCCGCTTCTA TCGAAGAAGC TTACAAGGCT GCACGCGCCC TCAATCCAAA CCTCATCATG 2340

CTTGGTGAAG GTTGGAGAAC CTATGCCGGT GATGAAAACA TGCCTACTAA AGCTGCTGAC 2400

CAAGATTGGA TGAAACATAC CGATACTGTC GCTGTCTTTT CAGATGACAT CCGTAACAAC 2460

CTCAAATCTG GTTATCCAAA CGAAGGTCAA CCTGCCTTTA TCACAGGTGG CAAGCGTGAT 2520

GTCAACACCA TCTTTAAAAA TCTCATTGCT CAACCAACTA ACTTTGAAGC TGACAGCCCT 2580

GGAGATGTCA TCCAATACAT CGCAGCCCAT GATAACTTGA CCCTCTTTGA CATCATTGCC 2640

CAGTCTATCA AAAAAGACCC AAGCAAGGCT GAGAACTATG CTGAAATCCA CCGTCGTTTA 2700

CGACTTGGAA ATCTCATGGT CTTGACAGCT CAAGGAACTC CATTTATCCA CTCCGGTCAG 2760

GAATATGGAC GTACTAAACA ATTCCGTGAC CCAGCCTACA AGACTCCAGT AGCAGAGGAT 2820

AAGGTTCCAA ACAAATCTCA CTTGTTGCGT GATAAGGACG GCAACCCATT TGACTATCCT 2880

TACTTCATCC ATGACTCTTA CGATTCTAGT GATGCAGTCA ACAAGTTTGA CTGGACTAAG 2940

GCTACAGATG GTAAAGCTTA TCCTGAAAAT GTCAAGAGCC GTGACTATAT GAAAGGTTTG 3000

ATTGCCCTTC GTCAATCTAC AGATGGCTTC CGACTTAAGA GTCTTCAAGA TATCAAAGAC 3060

CGTGTCCACC TCATCACTGT CCCAGGCCAA AATGGTGTGG AAAAAGAGGA TGTAGTGATT 3120

GGCTACCAAA TCACTGCTCC AAACGGCGAT ATCTACGCAG TCTTTGTCAA TGCGGATGAA 3180

AAAGCTCGCG AATTTAATTT GGGAACTGCC TTTGCACATC TAAGAAATGC GGAAGTTTTG 3240

GCAGATGAAA ACCAAGCAGG ACCAGTCGGA ATTGCCAACC CGAAAGGACT TGAATGGACT 3300

GAAAAAGGCT TGAAATTGAA TGCCCTTACA GCTACTGTTC TTCGAGTCTC TCAAAATGGA 3360

ACTAGCCATG AGTCAACTGC AGAAGAGAAA CCAGACTCAA CCCCTTCCAA GCCTGAACAT 3420

CAAAATGAAG CTTCTCACCC TGCACATCAA GACCCAGCTC CAGAAGCTAG ACCTGATTCT 3480

ACTAAACCAG ATGCCAAAGT AGCTGATGCG GAAAATAAAC CTAGCCAAGC TACAGCTGAT 3540

TCACAAGCTG AACAACCAGC ACAAGAAGCA CAAGCATCAT CTGTAAAAGA AGCGGTTCGA 3600

AACGAATCGG TAGAAAACTC TAGCAAGGAA AATATACCTG CAACCCCAGA TAAACAAGCT 3660

GAACTTCCAA ATACAGGAAT CAAAAACGAA AACAAACTCC TATTTGCAGG AATCAGCCTC 3720

CTTGCGCTCC TTGGTCTCGG TTTCTTACTA AAAAATAAAA AAGAGAACTA AACTAGCCCT 3780

CCTATAGAAA AATCCCCCAA GCATTATAGC TCGGGGGATT AATTTTTGTA CAATATTTGT 3840

TGTCCTAATA AACTTGATTA GGATTTTTTA TTAAGCCTCT TTCATAGCAA AATAAGCTCG 3900

TACTTTGGGT GCAACTTGTG TTCCGAAGAG TTCAATAGCT CTCAGAACCT GGTCATGAGG 3960

CATAGAACCA AGCGGTAGAT GAAGCATGAA GCGGTCCAAT CCTAAATCCT CTATCATGCG 4020

AATCAATTTT TCGGCCACCT GATCTGGATT GCCAACAAAC ATGGCGCCAT TTGGCCCTAC 4080

CTGCTCCAAA TATTGCTCAT AACGCAATTC CTGCCAGTGC GGACGGTCTT TGGAAATAGC 4140

ATCCACCACT TGCTTAGTCG GATGGAAATA ATCTTTCACC GCCTGCTCAC CATCTTCCGC 4200

AATCCACCCC CAAGAATGGG CTCCCACTTT CAAGTCTTTG TCAGCATGGC CCCTTCGCTT 4260

CCAATCTCAC GATAAGCCTG AATCAACTTT TTAAAATAAC GTGGATTACC ACCAATAATA 4320

GCATATACAA TCGGTAGACC AGCCTGAGCA ATCTTCACTG TTGATTCGAC ATGACCACCT 4380

GTAGCTATCC ACAAGGGCAA TTTGTCCTGA ACTGGACGAG GATAAACTTC TTTACCAGCA 4440

ATCGTTTGAG TCAATCGACC TTGCCAGTCT AACTTGGTCT TTTCATTGAC TAACTGAAGC 4500

AAGTCTAATT TCTCATCAAA AAGAGAGTCG TAGTCTTTCA AGTCATAACC AAACAGAGGG 4560

AAAGATTCCG TGAAAGAGCC CCTTCCAGCC ATAATCTCCG ATCGTCCATT TGACAAAGCA 4620

TCGATAGTGG CATACTGTTG GAACAAACGA ATCGGGTCCA TGCTTGACAG AATGCTGACT 4680

GCACTGGTCA AACGGATTTT CTTGGTATTG ACTGCCCCAG CGGCCAGAAC AATCTCTGGG 4740

GCTGATACTG CAAAATCCGC CCGATGGTGC TCACCAATCC CATATACATC CAAACCAACC 4800

TTGTCAGCCA GCTCAATCTC TGCCACCAAC TGGCGAATGC GTTCAGCATG ACTGTAAGTT 4860

TGTCCAGTCC CTTCAAGCTC CGTTATTTCC CCAAATGTTG AAATTCCCAA TTCTACCATT 4920

GTGATTCTCC TTATCTATCT CTGTACTTCA ATTTGAAAAA TTATTCTAAC ACGAATCTTG 4980

AGTACAAGCA ACCGATTTGC TCATTAGAAA AAGCCTAGAT AACTAGACTT TTTTAGCTTA 5040

TTCTACCGTT ACTGACTTGG CAAGGTTACG TGGTTTGTCC ACATCGAGGC CACGGTGGAG 5100

GGTTGCAAAG TAAGCGACTA ATTGCGTTGG TACGACCATT GAAATTGGTG AGAGGTATGG 5160

ATGTACGGTC GTAAGGACGA TATCGTCGGT ATCTTTGGCT ACATTCTCTT CTGCGATAGT 5220

GAGGACTTTG GCACCACGGG CTGCGACCTC TTGGATATTT CCACGAGTAT GATTGGCAAG 5280

AACTGGATCT GACAAGAGAG CCAAAACAGG CGTTCCTTCT TCAATCAAGG CAATGGTTCC 5340

GTGCTTGAGT TCTCCTGCAG CAAAGCCTTC ACACTGGATA TAAGAAATCT CTTTGAGTTT 5400

GAGACTTGCT TCCATGGCTA CGTAGTAATC TTGACCACGT CCGATGTAAA AGGCGTTACG 5460

AGTTGTTTCA AGAAGTTCAC GAACCTTGAC TTCAATGGTT TCTTTCTCTG AAAGAGTTGA 5520

TTCGATAGAC TGAGCTACGA TTGACAATTC ATGAACCAGG TCAAAGGCTT GCGCTTTAGC 5580

ATTACCATTT GCTTCTCCGA CTGCTTTTGC AAGGAAGGCA AGGGCTGCGA TTTGCGCTGT 5640

ATAGGCTTTA GTTGATGCCA CGGCAATTTC AGGACCTGCG TGAAGGAGCA TGGTATAGTT 5700

GGCTTCACGT GAGAGGGTTG AACCTGGAAC GTTTGTCACT GTTAAGCTTG GAATTCCCAT 5760

TTCATTAGCC TTGACCAAAA CTTGACGACT ATCCGCTGTT TCACCAGATT GGCTGATAAA 5820

GATGAAGAGT GGTTTCTTGC TGAGAAGTGG CATACCGTAG CCCCACTCAG ATGAGATTCC 5880

AAGTTCAACT GGTGTATCTG TCAATTCTTC CAACATTTTC TTAGAAGCAA ATCCTGCATG 5940

GTAAGATGTT CCAGCTGCAA GGATGTAGAT GCGGTCTGCG TCTTGAACAG CCTTAATGAT 6000

ATCTGGGTCT ACGACAACTT GACCAGCCTC ATCTGTGTAG GCTTGGATGA GTTTCCGCAT 6060

AACAGTTGGT TGCTCGTCAA TTTCCTTGAG CATGTAGTAA GGGTAAGTTC CCTTACCGAT 6120

ATCTGACAAG TCAAGTTCAG CAGTGTAGCT AGCACGCTCA CGACGATTTC CATCATAGTC 6180

TTGAACTTCC ACACTATCAG CCTTGACGAT TACCAACTCT TGGTCATGGA TTTCCATGTA 6240

TTGGTTAGTT TCACGAATCA TAGCCATGGC GTCTGAGCAG ACCATGTTAT AGCCTTCTCC 6300

AAGACCAATC AAAAGTGGTG ATTTATTTTT AGCTACGTAG ATGACTTCAG GATCTTGTGA 6360

GTCAACCAAG GCAAAGGCAT AAGAACCACG GATGATGTGA AGGGCTTTTT TGAAGGCTTC 6420

AAGAACTGAG AGCCCTTCTT CTTCCGCAAA TTTTCCAATC AAATGAACGG CTATTTCAGT 6480

ATCTGTCTGC CCCTTGAAGT GGTGACCTGC AAGGTATTCT TCCTTGATTT CAAGATAGTT 6540

CTCAATCACC CCATTATGCA CCAAGACAAA ACGTTCCGTC TCAGAGCGGT GTGGGTGAGC 6600

ATTGTCCTCA GTTGGTTTTC CGTGAGTAGC CCAACGAGTA TGTCCGATAC CAGTTGTTCC 6660

CTCAACACCA GCTGTCTTGG CAGACAATTC TGCAATACGA CCAACCGCCT TCACCAAATG 6720

GTTATCAGCA CCATCTAGGA CAAAAATTCC CGCAGAATCA TAGCCACGGT ATTCAAGCTT 6780

TTCAAGCCCT TGAATCAAAA TATCAGTTGC ATTTGTGTTT CCAACAACAC CAACAATTCC 6840

ACACATAGTA TATACGACAC AGGCAAGCTG TGCTTTCTCC TTAAAATTGG TATAGTCTAA 6900

TTCATCTTTT ATAGAATCAG CAAAAACAGT ATATACTTGT TTCTTTCACT TGTCAAGAGT 6960

AAAAATTGGT ATAGTTCAAA TTAAGCTCCT GTAAGCATAA AAACTCTGAC CGATTGGGAT 7020

AATCAGTCAG AGTCCTTTTT AAAATCCATT ATTATCGCTT AATTCTTTGA ACCAGTGGCC 7080

TGATTTCTTC AGACGACGTT CTTGCGTTTC CAAGTCTAAT TCGACCAAAC CATAGCGATT 7140

TTTATAGCTG TTGAGCCATG ACCAGCAGTC AATAAAGGTC CAAATCAAGT AGCCCTTACA 7200

GTTGGCACCA TCTTCAATGG CACGGTGAAG TTCACGAAGA TGACCTTTTA CAAAGTCAAT 7260

ACGGTAATCA TCTTGAATCA TTCCATCTTG ACGGAATTTT TCTTCCCCTT CAACACCCAT 7320

ACCATTCTCA GTCAACATCC ACTCAATATT GCCATAATTT TCCTTGATAT TTTGGGCGAT 7380

GTCATAAATC CCTTGCTCAT AAATCTCCCA ACCACGGTGA GAATTGATTT TACGTCCAGG 7440

CATCACATAA GGCTCGTAAA AATGTTCTGG TAAGAGTGGA CTCTCTGGAT GCTTAGCAAA 7500

TCGAGGAGCC ATAACACGCA AAGGTTGATA GTAGTTCACA CCAAGGAAGT CCACCGTATT 7560

ATCACGAATG AGTTCCAACT CTTCCTCTGT AGCATCAGGT AAAAGACCGT GTTCATGCAA 7620

GATTTCTACC AACTCCTGTG GATAAGTCCC CAAGACAGAT GGATCTAAGA AAGATTGGGC 7680

CTGAAAAAGG GCCGCAATAC GAGCTGCCTT GACATCAGCA GGATGCTGGC TACGTGGATA 7740

AGCCGGTGTC AAGTTGAGGA CAATCCCAAT CTTGGAATCA GGCAAAAGTT CATGGCAAGC 7800

CTTAACAGCC CGGCTGCTGG CCAATTGTGT ATGATAGGCT ACCTTAACAG CTGCCTCTGC 7860

ATCCACCTTA TGTGGATAAT GGGCATCATA AAAATAACCA AATTCTACAG GAACGATGGG 7920

CTCGTTAAAG GTAATCCATT GATCCACTAA ATCTCCATAA GTCTCAAAAC AAAAACGAGC 7980

ATAGTCTTCA TAGGCTGAGA CTGTCGCCTT ATTTTCCCAA CCATCACCAT CCTCTTGAAG 8040

GGCAAAAGGT AAATCAAAAT GATAGAGATT GACTAACAGA CGAATTCCTT TAGCCTTAAT 8100

AGCCTCAAAG ACCTTACGAT AAAAATCCAC ACCTTGAGTG TTGACTTTTC CACAGCCTTG 8160

TGGAAAAATC CGTGACCACT GAATAGAAGT CCGAAAGGCT GTGTGACCAG TCTCTAACAA 8220

AAGCTCAATA TCCCGCTCCC AATTTTCATA AAAAGTCGAT GTCTTATCTG AACCAATCCC 8280

ATTATAGTAA CGATTTGGCT CCACTTGGAA CCAGTAATCC CAGAGATTGT CTCCCTTACC 8340

GTCACCAGCT ACACGTCCTT CTGTCTGCGG TCCAGAAGTA GAGGATCCCC AGACAAAATC 8400

CTTTGGAAAT CTTAGCATAC ATTTACCTCT TTATCTACTC ATTTCTCCCA TTATACAGAA 8460

AAAACAAGGT AAAAACTAGT TACATTTTTT CCTTGTTTTT CTTCTGATTA TAGTTTTTAT 8520

TTCTTGCTTA GGATTTCAAG CGTTTCAAGC ACGTTATCTG CATGAACCTC AATGGTGTCA 8580

CCAGTTGCCT TGATCTTAAC TTCTACAATG CCATCGGCCG CTTTTTTACC AACAGTGATA 8640

CGGATTGGAA GACCAATCAA GTCACTATCG CTAAATTTAA CACCGACACG TTCGTTACGG 8700

TCATCTGTCA AGACTTCATA ACCAGCTCCC ATCAAGCTTG CTTCAAGTTT TTCTGTCAAG 8760

GCTTGCGCTT CTTCATCCTT GACATTGACA GTAATCAAAT GCACATCAAA TGGTGCCAAT 8820

TCTTTAGGGA AATTGATTCC CCAAGCGTAA CGGTATTCAC CTTTTGGCGT TTTGTTAACA 8880

AAGAGGCGAG CGTGTTGCTC CATCACTGCT GAAAGAAGAC GGCTGACACC GATACCGTAA 8940

CATCCCATGA TGATTGGCAC AGCACGACCA TTTTCATCCA AGACATCTGC TCCCATGCTT 9000

GCTGAATAGC GAGTTCCGAG TTTGAAAATA TGACCGATCT CAATACCACG CGCAAAGTTA 9060

AGGACACCTT GTCCATCTGG GGAAATTTCA CCCTCACGAA CTTCACGGAT ATCCACATAT 9120

TCTGCAGTAA AATCACGGCC TGGGTTCACA CCAGTCAAGT GGTAGTCATC TTCGTTAGCA 9180

CCGACAACTG CATTGCGAAC ATCTTGTACC TTACGATCTG CAATAATTTT AATATTCTCT 9240

GGCAAACCAA CTGGTCCAAG TGAACCAAAT CCTGCTTGAA CAACATTCGC CACTTCTTCT 9300

TCGCTAGCAA CGTCAAAGAA ATCTGCTCCC AAGTGATTTT TCAACTTGAC TTCGTTGAGT 9360

TGGTCATTTC CAACTAGAAG GGCTGCAACA AGCTCACCAT CTGCAATGTA GAAGAGGGTT 9420

TTAATCGTTT GTTCTTCTGG AACATTGAGG AAGGCTGCAA CTTCATCAAT TGATTTAACA 9480

TCTGGCGTTG CAACACGAGT AACTTCTTCT TCAGCGACAA CACGGTTGCT TGGTTTGTAC 9540

TCGTTTGTTG CCATTTCTAA GTTAGCTGCA TAGCTAGACT CACTTGAGTA AGCAATGGTA 9600

TCTTCACCAG AGACTATCCA TTTGAGCAAT TCTGCCTTGA TTTCTTCTTG CACTTCTGCA 9660

GGAATTTCGT CAAATGAGGC AACTGACTTG TCCAAGACAA CCCAGCGGTC AAGGTCTGTA 9720

CGAGCAGATG TAATGGCCAT AAATTCTTGG CTATCCTTAC CACCCATGGC TCCACCGTCA 9780

CCAATAATAG CCTTGAAGTC TAAACCACTA CGAGTGAAAA TACGCTCATA GGCTGCTTTG 9840

TACTCATCAT AAACACTATC CAAACTATCA TAGTTAGCGT GGAAACTATA AGCATCCTTC 9900

ATGATAAACT CACGTGTACG AAGAAGTCCA TTACGCGGGC GTTTTTCATC ACGATACTTG 9960

GGCTGAATTT GATAAAGGTT GAGTGGCAAT TGCTTGTAAG ATTTAACAGA ATCACGGACA 10020

ATAGCTGTAA AGGTTTCTTC GTGAGTTGGA CCTAAGATAA AGTCTGATTT TTCACGGTTT 10080

TTTAGTTTGT AAAGGTCTTC ACCATAGGTT TCGTAACGAC CTGATTCACG CCACAATTCT 10140

GCACTAAGAA GGGCTGGAGC CAACATCTCA ACAGCACCAA TCTTTTCGAA TTCTTGGCGC 10200

ATGATGTTTT TAGCTTTTTC AATCACACGG TTGGCAAGTG GTAGATAAGA ATAAACACCT 10260

GCTGAAACTT GGCGAACATA ACCAGCACGC AACATAAGAG CATGGCTGAT AACTTGAGCA 10320

TCGCTTGGCA TTTCGCGAAG CGTTGGGATA GGCATTTTAC TTTGTTTCAT AATATTCCTC 10380

GATTATCTAA AAAAGAGTCG CATAATGTCA TTCCAAGTCA CAGCAATCAT CAAGACAACC 10440

ATGATGACCA CTCCGGCCAA GGTGACATAG GTTTCAATTT CTTGTTTCAA TGGTTTGCCG 10500

CGGATGGCTT CTAGGATATT GAGCACAATC TTACCACCAT CCAAGGCTGG AATCGGAATA 10560

AGATTAAAAA TCCCAATATT GATGGAAATC ATTGCCAAGA AGTACAAGAT ATTTTCAATT 10620

CCATTTTTAG CAGCATCACT ACTTGCCTTA AAGATAGCAA CAGGTCCACC CAACTTGTTC 10680

AAATCTGGTT GGAAAATCAG ATTTTTCAGA GCTGAGAGAA TTCGGAGAGC TGAGTCAGCA 10740

GCAGTTGTAA AACCACCTAC AAACATGGAT AGAAAATCTG ACTTAACCCC CGGTTGAACA 10800

CCTAGAAGGT AACGACCTTG ACTATCTTTG GGTGTAACAG TGACTTGTTT GTCACTCCCC 10860

TTTTCAGAAA TAGTCACATC CAAAGTCGGT GCCGTCTTAT CTTTGGTTTC TGTTTCCACA 10920

GCTTGGATCA AGCTTTCCCA GTTGCTAACC TCATGTGAGC CAATCTTGGT AATTTGTGCC 10980

ATTTCTGGTA CTCCTACCTT GGCCAAGGCA CCTTGGGGCA TGATATGGAA CTGATTGGTA 11040

TCAACATCTC TGACACCACC CTGCATAAAG ATTAAAACCC AAAAAACAAC GACACCTAAG 11100

ATAAAATTGT TCATAGGACC TGCAAAATTG GTAATCAGTT TGCCCCAGAT AGTCGCATTT 11160

TGATATTGAA CATCTAAAGG TGCAATCCGA ACCTCAGTAC CATCTGCTTC CACAACCGTT 11220

GCATCGTGAT CCACTGCAAA TGTTTTTTCT TCTTCCAGAA CCAATCCTTT GATAAAGAGC 11280

TTGTCTTCAA AATCAAACTG GGTCACCTGC ATAGGGAGGG CTGTTTGATC CAATTTTTTA 11340

CCTGAGAGAT TGATGCGTTT AACCTTACCA TCATCAGCAA GTGTCAAACT AACAGGCGTT 11400

CCTGTCTTGA TTTCAGTTGT ATCATCACCC CAACCGGCCA TGCGGACATA GCCACCCAGA 11460

GGCAAGATTC GAATGGTATA GGCCGTTCCA TCCTTGCCAA TGTGAGCAAA AATTTTAGGT 11520

CCCATACCGA TGGCAAATTC ACGTACTAAA ATCCCTGATT TCTTGGCAAA GTAGAAGTGA 11580

CCGAACTCGT GCACCACTAC AATAATCCCG AAAACCAGAA TAAAGGTTAA AATTCCGAGC 11640

ATAGCGTTTC CTCCGTCTTT TGATTAAAAG AGTCCAAATA AGTGCATGAT TGGAAATACA 11700

AGCAACATAC TATCGAAACG ATCCAAAACA CCACCATGTC CAGGGATAAA TTTCCCAGAA 11760

TCCTTAACAC CAAAATGACG TTTGATCGAA CTTTCTAGTA AATCACCAAA TTGTCCAGCA 11820

ATGCTAAAGA AAATAGCAAA GACTGACATC TTGTAAATTC CATATGGAAG AGCAACTGTA 11880

CTGTCAACTA TCATAAGGAT AATGGTTACT AAAATTGCTC CTAAAATACC ACCCAAGGCA 11940

CCCTCAAGGG TTTTATTAGG CGATACCCTT GGTGCTAACT TTCGTTTCCC ATAGTTCATC 12000

CCAACAAGAT AGGCACCACT GTCTGTCGCC CAGACGATAC ACAAGGCTAA GAGAGCCTTG 12060

TCCAAACCTG CAACACGAGC ATCTAGTAAA GCATTAAATC CAAAGCCCAC GTAGAAGCTC 12120

ATAGCAAGAG GGAAAACCGC ATCCTCAATC GTATAAGACT TGCTAAAAAC GGTCGTTCCT 12180

AACATGATTG AAATCAAAAC ACTATAGGCA ACCACATTCC CATCAACTGG CAAAAAAGTC 12240

AGGTAATTCT CCAAGGGAAT GGTCAATGCA AAGGTTGCAA AGAGGGTCAA GAGGCCCTCC 12300

ATCGTCATGG TCTCTAGACC TCTCATCTTC AAAAGTTCAT GCATGGCTAG CATGGCTATG 12360

ATTCCGATTG CTATCTGAAG CAAGAGGCCC CCAATCATTA AAATTGGTAG GAAAATAGCC 12420

AGGGCAATCC CTGCAAACAA GGTTCTTTTC TGTAAATCCT GGGTCATATT TCCTCCTAAA 12480

CTCCTCCAAA TCGGCGATGA CGACGATTAT AGGCAAGAAT AGCTTCCTGC AAGGCCGCTT 12540

CGTCAAAATC AGGCCATAAG GTGTCCGTAA AATAAAGCTC ACTATAGGCT CCCTGCCATG 12600

GAAGGAAATT GCTCAAACGT AATTCTCCAC TAGTACGGAT AATCAAGTCT GGGTCTCGTA 12660

AGTCCTTAGG CAAATGCTGA GTAAAGAGAT AGTTACCAAT CAATTCCTCT GTGATGTCAC 12720

CTGGGTTGAT TTTGGCATCT AAAACATCCT GGGAAATCAA CTTAAGCGCC TGTGTAATCT 12780

CAGCACGTCC ACCATAGTTA AGAGCAAAAT TAAGAATCAA TCCTGTGTTG TTCTTAGTCA 12840

ATTCCTCAGC CTTGGTTAAA GCTTCAAAGG TTTGCTTAGG CAGGCGGTCT GTCTCCCCAA 12900

TCATTTGAAT CTTAACATTA TTCGCATGTA GTTCCGGGAC ATAATTATCA TAAAACTCTA 12960

CTGGCAAGTT CATGATAAAC TTGACTTCCT GATCTGGACG GGTCCAGTTT TCCGTAGAAA 13020

AAGCATAGAC CGTAATAACC TTGACGCCCA GTTTGTTGGC TGCCTTGGTC ACGGTTTGCA 13080

ATGCTTCCAT GCCCGCCTTA TGTCCAAAAA CTCGCGGTTG CATACGTTTT TTAGCCCAAC 13140

GGCCATTGCC ATCCATGATG ATGCCGATAT GAGCAGGAAC CTGTGTCGGA ACCTCTACTT 13200

CCACAGCCTT ATCTTTCTTA AAAAATCCAA ACATGATCTT ATTCCTATTC AAAAATCTAT 13260

CGTTTCATTA TACCATATTT CCCCATTTTC TTCTATCACT AAGCTATTTA TTCTCAGGCA 13320

CCAAGCCCAT TTTTCAAAAA AATAAGCCGC CTGATTGGGC GACTTTATTT TTATAGGGAG 13380

ATTATTATGA AAAAGTTTTA GGAGTTTAAG TTAAGGTCTT CTTAACTTAT GAACTTAGTG 13440

TACACTCCCT AGCTTAAAGT TTCCTTAAGT ATTTTTAAAA ATCAAATTTT TCCATTTCTC 13500

CTGCCAATTT TTCTTGGATA AACGTGTTTG ATAGAGTTCC ATTCGGTCTT CATTTTCTAA 13560

GAAATGAGGA GTTGGACGAA CTTGAAAATT CAAAATATCC TCCAAACCAT AAGGTACATA 13620

GAGTTCAAAA TCTAATTCTT CATTCAAGCG CAGTCCAACT GCCGTACACC GTTCTGGATA 13680

CTTACTCATA GCATCACGAG AACTGGTATA GGAAGCAGTG TGAGGACTGT GCTGATGCAT 13740

ATAGACCTGA TTTTTCAATT CCCACTGGTA CTGAGGAAAA TCCTCTCTCA GCTTTTTCTC 13800

CAGTAATAAG GTTTCCTCAT AAGAAAAATC TGGATCAAAG AAAATCACAT CTATATCTGT 13860

TTCATGATCA AAAGGGGATT TGTCTGACAA AAGATTCCAG ATGAAATTTC TGACAGAACC 13920

TGCTGCCAAC CACGAGTCTT TCAAACCAAG GTCTCGGATG ATCGTCAGAA TGGCCATCAT 13980

ATCTGGACTT TCTCTAAAAG CCTCTAAGAT TTCTTGCTTA TTTTTCACTG TATTCATAAC 14040

CTAAGTGCTC ATATGCCTTA GCAGTCGCCA CCCGTCCAGA CCGTGTCCGC ATGATAAAAC 14100

CTTTTTGAAT CAAGTAAGGC TCATACATGT CTTCAACTGT CTCACGCTCT TCGGCGATAT 14160

TCACAGAAAG AGTTCCTAGA CCAACAGGTC CTCCACTGTA CATCTCAATC ATGGTGCGAA 14220

GGATTTTTTG ATCCACATAG TCCAAACCTT CATGGTCAAC ATCCAGCATA GTCAAAGCCT 14280

TATCGGTAAT AACATCATCG ATAACCCCAT TCCCCATTAT CTGGGCAAAA TCGCGCACGC 14340

GCTTGAGGAG ACGATTGGCA ATACGAGGGG TTCCACGACT ACGTAGGGCC AACTCAGATG 14400

CTGCCTCATG GGTGATTTCC ATCTCAAAAA TATCTGCCGT CCGCTCGACA ATTTCTGTCA 14460

AGTCAGCATG AGCATAATAC TCCATATGAC CTGTAATCCC AAAACGTGCC CGTAGTGGAT 14520

TTGAGAGCAT ACCAGCCCGA GTCGTCGCAC CAATCAAGGT AAAAGGAGGC AACTCCAAAT 14580

GAACACTGCG ACTGCCTTCA CCAGCCCCAA TCATAATATC GATGTAGAAG TCCTCCATGG 14640

CACTATAAAG CACTTCTTCC ACTGACATGG GTAAGCGATG AATCTCGTCA ATAAAGAGGA 14700

CATCTCCAGG CTCTAAATCA TTCAAAATCG CTACCAAATC ACCCGCTTTT TCGATAACAG 14760

GACCAGACGT TTGCTTGAGA TTGACTCCCA GTTCATTGGC AATGACAAAA GCCATGGTTG 14820

TTTTCCCAAG CCCTGGAGGG CCAAATAAGA GCACATGATC CAGCGCTTCA TCCCGCATTT 14880

TAGCGGCTTC GATAAAGATC TGAAGTTGAT CCTTAACCTT ATCCTGACCA ATATATTCAC 14940

GTAAATACTG AGGACGGAGC GTGCGTTCTA CTAACTCCTC ATCACCCATC ATCTCATTAT 15000

CTAAAATTCT ACTCATGGCT CTATTATATC AAAAAAAACA AGCCACAAAC AAAAAAGCCA 15060

CCTGATTGGG TGACTCCTAA GTTTAGCACT TATGTGGTAT AATATTATAC GGCACTTCTA 15120

CACCGCCTAC GAAAGGAGGT GAGATAGCCC ATGATGGAAT TAGTACTCAA AACTATTATC 15180

GGACCAATTG TGGTCGGTGT CGTTCTTCGT ATAGTCGATA AATGGCTAAA CAAGGACAAA 15240

TAGTGTCAAA AAAGACCTCA AGCTTATTTG GTCGTGAGCT TGGGGTCTTT TCTAGCCTAT 15300

GATATAGAAC TAGTACTCAA TTCCTTTTTA TTATCCCATA GTTCACGAAT TTTGTCAAAA 15360

CTTTACATTT TCTTCAACCG CTGTACGACA AGACGGTTAA GATTAAGAGA ACGTTAGGGA 15420

TTCTATCAAT TTCATAGAAA TTTTGATTTC GTAAACGAAG AGACAATCTT ACATGTCACT 15480

TCTCATTTAA TACGCCACTA CTAGACAAGC AAAATCATTA TTACAGTAGT TCCAGTCCTT 15540

CAATTAACAG TCACTTACAA TCAAATTGAG TTTGAACTAG CTGAAGCGAC CACAGACCTA 15600

TTTCTTAGTC ATATTCGCTA AAAAAATCCC CGCCAAAATC TCAAAAAGTC CCCGCCAATT 15660

CCCCGACCAA AATCCGAAAA ATACCGAAAA ATATCGAAAA ATTATTTTTA GAATAGTCCC 15720

AAAAATCCTG AAATAGAGCT AAAAAACTCC ACCTGATTCG GTGGAGTTAA GGGAGATTAT 15780

TATGAAAAAG AAAAGTTTAG GATTTTATTA AATAAAGTTA GGAGGTCTTT ATTTAATAAC 15840

TACATGATAC AAGACGAAAC TTAAAACTAG CTTAACTTTT CTAAAATTTT ACTATTTTGC 15900

AAAAAATTTC TATCACCAGC ACCTCACCAA TCGAGTAGGG GATAATCTCT AGCCCCTCTC 15960

ACACCACCGT ACGTGCCGTT TGGCATACGG CGGTTCAACT AACTTTTAAC GCATGTCGTT 16020

CAAGGTAATA ATCCAAACAC GAAACCAGTC CACGTTTTTC CAGGACTGGT TTTGATATAG 16080

CACGTTTAAG TACCGACTTC TGAGCTACTA ATTGATAATG GTCGCCCCAG CCAGATACCT 16140

TATCTGCTAT CCATTTAGGA ACTCCTAACT TAAGCAATCC CCATAATCGT CTCGATTTCT 16200

TCTTCCATTG CTTCCAGATA ATCACTCGTA GGCGAGTACG CAAGCGCTCA TCTATGCTGG 16260

CGACTATACT TTTCATATTT CCCAATGAGC AATAGTTTAT CCATCCTCGA ATAGACAAAT 16320

TCAGTTGCTC AATACGTCTT GTTAGGTCTA TACTCCATTT CCTCTGTGTT AGTTTCTTCA 16380

ATTTAAACTT AAATCTCCGA ACACTATCTT GATGTGGACG GCTTTTCCAA CCATCTGATA 16440

ATTTCCAGAA CCCAAAACCT AGATATTTCA ACTCTCTTGG TCATGTTTAC TTTCAAACCT 16500

AGCCGTTTCT CAATAAACGA CTGACTGAAT ACATC 16535

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8136 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

CCAGAGCGTT GCGTCCGAAA GTCTATCCAG ACACGGCTCT TTAAAAACAA AAGGAGAAAT 60

GATGCATACT TATTTGCAAA AGAAAATTGA AAATATCAAA ACAACCCTAG GTGAAATGTC 120

AGGTGGTTAC CGTCGTATGG TTGCGGCTAT GGCTGATTTA GGATTTTCAG GAACTATGAA 180

GGCTATCTGG GATGACCTCT TTGCCCATCG TAGTTTTGCC CAGTGGATTT ATTTGCTGGT 240

TTTAGGAAGT TTTCCTCTCT GGCTGGAGTT GGTTTACGAA CATCGTATTG TTGACTGGAT 300

TGGGATGATT TGTAGCTTGA CAGGGATTAT CTGTGTAATC TTTGTATCGG AAGGTCGAGC 360

AAGTAATTAT CTTTTTGGCT TGATTAACTC TGTTATTTAC CTTATTTTGG CCCTACAGAA 420

AGGCTTTTAT GGTGAGGTGC TGACGACACT TTACTTCACA GTCATGCAGC CAATTGGACT 480

TCTAGTTTGG ATTTATCAGG CACAGTTTAA GAAGGAAAAG CAGGAGTTTG TCGCGCGTAA 540

ACTGGACGGC AAGGGCTGGA CAAAGTATCT TTCCATTAGT GTGCTTTGGT GGTTGGCCTT 600

TGGCTTCATT TATCAGTCTA TTGGTGCCAA TCGTCCCTAT CGTGATTCAA TCACAGATGC 660

AACCAATGGG GTAGGGCAAA TCCTCATGAC AGCTGTTTAC CGTGAACAGT GGATATTCTG 720

GGCGGCTACC AATGTCTTTT CAATCTATCT CTGGTGGGGA GAAAGCCTGC AAATTCAAGG 780

GAAATATCTA ATTTATCTCA TTAACAGTCT AGTTGGTTGG TATCAATGGA GCAAGGCAGC 840

TAAGCAGAAT ACTGATTTAC TTAACTAGGA AAAGATGTTT GAAAGTGCTG TTTTGAGATT 900

TCGATTAAAA CAGATATAGT TGATAATCAA GGATTTATAG TATGAAAAAG AGGATCGGCG 960

GGTCCTCTTT TGTTGTTGAA AAGATAAAAA ACTCAGTAAC CTAGAAATAA GACAACTGAA 1020

GCTTTACTCT ATATTCAATT TTTAGGAATG AGAAGGTCTA GATAAAATTG GACAACTTCC 1080

TGGTCTGTGA AATCTTGACC TTTTTTGAGC CACCAGGTCA ATGTCTCGAT AAAGTTGGAC 1140

ATGACCAAGT GTTGGAGGTA AGAAGTAGGC AGATTAGGGT GGGCTTCTTT TAAATTATCA 1200

GCTAGCACGG AATAGACATG GTGTTCTAGC TCTTTATGGA GTTGACGGAG GAAGTAGTCA 1260

TTTTTGGAAA ATAGCAGACT GGTGATATGG TCTTGGTTTT TATGAAAATG GAGAAAGAGG 1320

TGGGCGAGGT AGTCCTCGGT TGAAATGGCT TGCTCTCTTT CAAAAAGATG ATGGAAGAGG 1380

TAGCGGCAGA GCTGGTCCAG AAGAAGCTCC TTACTCTCAT AGTGACAGTA AAAGGTGGAT 1440

CGTCCCACAT CTGCGAGATC AATGATATCC TGAACAGTAG TGGCCTCGTA GCCCTTAGCA 1500

TTCAAAAGTT GTATAAAAGC TTGATAGATG GCTTTTTTGG TTTTGCTGAT ACGGCGGTCA 1560

ATGTTAGTCA TATGGACACT TAAGGCAAAT TGTTCAGAAC TGAATAAAGC TGACGTTTTG 1620

CTTCTATCCT TTCTTTGAGT TTTAGTGGAT AATGATAATG AACAAGGTGT TCATAAATCT 1680

ATTATAACAA AGGAATGAGA AATATGAAGG CAAAATATGC TGTTTGGGTG GCTTTTTTCT 1740

TAAATTTGAC TTATGCCATT GTTGAGTTTA TTGCAGGTGG AGTATTTGGT TCTAGCGCTG 1800

TTCTTGCTGA CTCTGTGCAT GACTTGGGAG ATGCGATTGC AATTGGAATA TCAGCTTTTC 1860

TAGAAACAAT CTCCAATCGT GAAGAAGACA ATCAGTACAC CTTGGGCTAT AAGCGGTTTA 1920

GCCTGCTAGG AGCCTTGGTA ACAGCTGTGA TTCTCGTAAC GGGCTCTGTT CTAGTCATTT 1980

TGGAAAATGT CACGAAGATT TTGCATCCGC AACCAGTCAA TGATGAGGGG ATTCTCTGGT 2040

TAGGAATTAT TGCGATTACT ATCAATCTGT TAGCGAGTCT GGTGGTTGGT AAGGGAAAGA 2100

CAAAGAATGA GTCTATTCTG AGTCTGCATT TTCTGGAAGA TACGCTAGGG TGGGTAGCTG 2160

TTATCCTGAT GGCGATTGTT CTTCGATTTA CGGACTGGTA TATCCTAGAT CCTCTTTTGT 2220

CCCTTGTCAT TTCTTTCTTT ATTCTTTCAA AAGCCCTTCC ACGTTTTTGG TCTACACTCA 2280

AGATTTTCTT GGATGCTGTG CCAGAAGGTC TTGATATCAA GCAAGTAAAG AGTGGCCTGG 2340

AGCGATTGGA CAATGTGGCC AGCCTTAATC AGCTTAATCT CTGGACTATG GATGCTTTGG 2400

AAAAAAATGC CATTGTCCAT GTTTGTCTAA AAGAAATGGA ACATATGGAA ACTTGTAAAG 2460

AGTCTATTCG AATTTTCCTA AAAGATTGTG GTTTTCAAAA TATTACCATT GAAATTGATG 2520

CTGACCTAGA AACTCACCAA ACCCATAAGC GAAAGGTGTG TGACTTGGAA CGGAGTTATG 2580

AGCATCAACA TTAGAAAAAA GTGAAAAATA CTTGGGTACT ATCTTATTTG GAATAGAGTA 2640

ATTTCTTTAT TATTTAAATA TTTCAAAAAT TGGTAAGAGA AGAGCATTGT ATAAACTCCA 2700

GATATATGAT TGTTAATGAT AAAAATTTTT CGATTAGATA CAAAATGCTT GACTTGGAGT 2760

CAACTCAAAG TTATATAATA AGATAAGTGA GTTAGAATAG CGTGAATTCA GTGAATGAAA 2820

TGAGAGGAGG TTAGCGTGTG AATATTAAAT CTGCCAGTGA TTTGTTGGGA ATTTCAGCGG 2880

ATACGATTCG GTATTATGAA CGGGTTGGTC TTGTGCCACC GATTACTCGT ACTGCTACTG 2940

GGATTCGTGA TTTTCAAGAT CAGGATATCG AAGCGCTGGA ATTTATTAAG TGTTTTCGTT 3000

CGGCGGGTGT CTCTGTAGAT AGTTTAGTTG ACTATATGTC GCTCTACCAA AAGGGAGATG 3060

AAACGAGAGA GGAGAGGCTT GGTATTTTAG AAGAGGAAAA GCAAAAATTA GAGGAGCGCT 3120

TGTCTCAGCT ACAGACAGCT TTAAATCGTT TAAATCTCAA AATTAAACTT TATAAGGAAG 3180

GAAAATTTTA AATGAAATCA GCAGTATATA CAAAGGCAGG TCAGGTTGGA CTTGCTAGCA 3240

TTGAACGTCC GCAAATAATA GAAGCGGATG ATGTGATTAT TCGTGTGGTT CGTGCGTGCG 3300

TTTGTGGTTC AGATTTATGG AGGTACCGTA ATCCAGAAAC GAAAGCTGGA CACAAAAATA 3360

GTGGACACGA AGCGATTGGG ATTGTTGAAG AAGCTGGGGA AGCCATTACG ACGGTGAAAG 3420

CAGGTGATTT TGTGATTGTC CCTTTTACAC ATGGATGTGG TGAGTGTGAT GCCTGTCTTG 3480

CTGGATTTGA CGGTTCTTGC GACAATCATA TTGGCAATAA TTTGGGGGGT GATTTTCAGG 3540

CAGAATATAT TCGCTTCCAC TATGCAAACT GGGCGCTGGT TAAAATCCCT GGTCAACCTT 3600

CTGACTATAC AGAAGGGATG CTCAAGTCCC TTTTGACTCT TGCAGATGTC ATGCCGACAG 3660

620

GCTATCATGC GGCGCGTGTT GCAAATGTTC AAAAAGGGGA CAAGGTTGTT GTTATCGGTG 3720

ATGGGGCTGT TGGTCAATGT GCTGTCATCG CGGCTAAGAT GCGTGGAGCA TCACAAATTA 3780

TCCTTATGAG CCGTCATGAA GACCGTCAAA AGATGGCTAT GGAGTCAGGT GCGACAgcTG 3840

TTGTTGCAGA ACGTGGTCAA GAAGGAATTA CCAAGGTGCG TGAAATCCTC GGTGGAGGAG 3900

CAGATGCAGC ACTTGAATGT GTTGGTACGG AGGCTGCTAT AGAACAGGCG CTAGGTGTTC 3960

TTCATAATGG AGGGCGTATG GGCTTTGTAG GAGTCCCACA CTATAATAAT CGTGCTCTTG 4020

GTTCGACATT TATGCAAAAT ATCTCTGTAG CAGGTGGGGC AGCTTCTGCT ACAACATACG 4080

ATAAGCAATT TTTACTAAAA GCCGTCCTTG ATGGTGATAT CAATCCAGGT CGCGTCTTTA 4140

CTTCAAGTTA TAAACTGGAA GATATCGACC AAGCCTATAA AGATATGGAT GAACGTAAGA 4200

CAATTAAGTC TATGATTGTA ATCGAATAAA AAACGAATAG GAGTTTTAGA ACTCTATTCG 4260

TTTTTTATGT TATCCTATTC TTGATTTAGG GTACTTTCTC TTAATGTCAG TCTGGTTCCC 4320

AGCATGGTCA GGCTAGGGAT TTTCCGACCG TGGAGGACTT CCTTGTTAAG AATATCCATA 4380

CCTGCTCGGC CCATTTCTTC AGTATAAACT GTAATACTAG AGAGGGGAGG ATAGACCTGT 4440

TTGGTCAGAC TAGTGTCGTT AAAGGAAATG AGGCTGACGC GATCTGGCAG GCTGATTCCA 4500

GCTTCTTGGA GGGCACGGAG GGCACCGATA GCTAAACTAT CGCTGGCTGC GAAAAATGCT 4560

GGCGGAAGTT GGTCTCCCAA GCTCTGAATG GCCTCCTTCA TTAAGTCATA GCCAGACTGG 4620

GCAGTAAATC TTCCTTGAAA GACCAGTTCA TCATGATAGA TTCCCCTCGC TTGACTATAG 4680

TTTTTGAAGT TTTCTAGACG CTTGTCCTGA ATGATTTCTT CTTGGTCTGT TGTTTCTTCA 4740

AGGCCTGTTA GAATCCCGAT ACGGTCCATT CCTTGACTGA GGAAATAATC GACAACCTGT 4800

TTCATAGCAG TGTAAAAATC CGTGATAATA CAGGTATGTC CCAGGGAAAG TGTATCGCTG 4860

TCTAGAAATA CAAGAGGCTT TTGGTATTCT TCAAAGGCAG AAATCTGAGC TCGACTAAAC 4920

TTTCCGATGC AGAGAATCCC AATCACTTCC TCGCTTAGGG TAAAAGGGTG GTCATTAAAA 4980

TAGCGCAAGA TATCATAGTC CAACTCTTGG GCTCTTTTTT CTATTCCTAG GCGAATCTGG 5040

TAGTAGTAGA GGTCGTCCAG CTCCCCTTGT TCGCTGACCC ATTGGATAAT GGCAATCTTT 5100

TGCTTGGGTT TGTGGGACTC GCCTGTCTTG AGGTGCTTGG TGTAGCCCAG CTCTTCAGCA 5160

ACGGTTAAAA TACGGTGTCT GGTTTCTTCT GTAACAGATA GGCTCTGGTC GCGGTTGAGG 5220

ACGCGGGATA CGGTCGCGAT AGAGACAGAG GCTAGCTGTG CAATGTCTTT TAAGGTAGCC 5280

ATAAATCCTC CTTGATTAGG TTAGTATATC ATGTTTTTCT TCTTTTTACT GATATTTTAC 5340

TAAAATTTTA GTAAAAAGGA TTGACCTTGG AAAATTCCTT GGATATAATA GAAAGAAAAC 5400

GATTACACGT TAAGATGGCT TAACGGACAG TCAAAGGAGA ATTCATATGG CACAACATCT 5460

TACTACTGAA GCCCTTCGCA AAGACTTTCT TGCTGTTTTT GGTCAAGAAG CAGATCAAAC 5520

CTTCTTTTCA CCAGGCCGCA TTAATTTGAT TGGTGAACAC ACAGACTACA ACGGTGGGCA 5580

CGTTTTTCCT GCTGCTATTT CCTTGGGAAC TTACGGTGCA GCTCGTAAGC GTGACGACCA 5640

AGTCTTGCGT TTCTACTCAG CTAACTTTGA GGACAAGGGC ATTATCGAAG TGCCTCTCGC 5700

TGACCTCAAG TTTGAAAAAG AGCACAACTG GACCAATTAT CCAAAAGGTG TCCTTCATTT 5760

CTTGCAAGAA GCTGGGCACG TGATTGACAA AGGTTTTGAT TTTTATGTTT ATGGAAATAT 5820

TCCAAATGGT GCTGGCTTGT CTTCTTCTGC ATCCTTGGAA CTCTTGACAG GAGTCGTGGC 5880

TGAGCATCTC TTTGATTTAA AATTAGAGCG TCTCGATTTG GTTAAAATCG GCAAACAAAC 5940

AGAAAACAAC TTTATCGGAG TAAACTCTGG CATTATGGAC CAGTTTGCTA TTGGTATGGG 6000

GGCAGACCAA CGTGCTATTT ACCTAGATAC TAATACTTTA GAATACGACT TGGTGCCACT 6060

TGATTTGAAG GACAATGTCG TTGTTATCAT GAACACCAAC AAACGCCGTG AATTGGCGGA 6120

CTCTAAATAC AATGAACGTC GTGCTGAGTG TGAAAAAGCA GTGGAAGAAT TGCAAGTTTC 6180

CTTGGATATT CAGACTCTGG GTGAATTGGA CGAGTGGGCC GTTGACCAAT ATAGCTATCT 6240

GATTAAAGAT GAAAATCGTT TGAAACGTGC TCGCCATGCT GTGCTTGAAA ACCAACGTAC 6300

CCTCAAAGCT CAAGTAGCAC TCCAAGCAGG AGATTTGGAA ACATTTGGAC GCTTGATGAA 6360

TGCGTCACAC GTTTCTCTGG AGCATGATTA TGAAGTAACT GGTTTGGAAT TGGATACCCT 6420

TGTTCACACA GCTTGGGCAC AAGAAGGAGT TCTCGGTGCT CGTATGACAG GGGCTGGTTT 6480

TGGTGGCTGT GCCATTGCCT TGGTTCAAAA AGATACTGTT GAGGCCTTTA AGGAAGCTGT 6540

AGGCAAACAC TACGAGGAAG TAGTTGGATA CGCTCCAAGC TTCTATATCG CTGAAGTTGC 6600

AGGTGGCACT CGCGTCCTTG ACTAGTCAAA AGGAGGCTCT ATAGTGACCT TAGTAAATAA 6660

ATTTGTAACA CATGTCATTT CTGAAAGCTC ATTTGAGGAA ATGGATCGAA TCTATCTGAC 6720

CAATCGTGTT TTGGCACGAG TGGGAGAAGG TGTTTTGGAA GTTGAGACCA ATCTGGATAA 6780

ATTGATTGAC CTCAAGGACC AGCTGGTTGA AGAAGCCGTT CGATTAGAGA CGATTGAGGA 6840

TAGTCAGACT GCGCGTGAAA TCCTTGGTGC TGAACTGATG GATTTGGTGA CTCCTTGTCC 6900

AAGTCAGGTC AATCGTGATT TTTGGGCAAC CTACGCCCAC TCTCCAGAAC AAGCGATAGA 6960

GGATTTTTAC CAACTCAGTC AGAAAAATGA CTACATCAAA CTCAAGGCCA TTGCTAGAAA 7020

TATCGCTTAT CGTGTTCCAT CTGACTACGG AGAACTTGAA ATTACCATCA ATCTCTCTAA 7080

GCCTGAAAAA GATCCCAAAG AGATTGTGGC AGCCAAGTTG GTGCAAGCTA GTAATTATCC 7140

TCAGTGTCAG CTTTGTCTAG AGAATGAGGG CTACCATGGT CGAGTTAACC ACCCAGCTCG 7200

TAGCAATCAC CGTATTATCC GTTTTGAAAT GGTTGGTCAG GAATGGGGTT TCCAGTATTC 7260

GCCCTATGCT TACTTTAATG AGCATTGTAT CTTTTTAGAT GGCCAGCATC GTCCCATGGC 7320

CATTAGTCGT CAGAGTTTTG AACGTCTGTT GGCTATCGTA GACCAGTTTC CAGGATATTT 7380

TGCTGGATCT AATGCCGACC TGCCGATTGT GGGGGGCTCT ATTCTAACTC ATGATCATTA 7440

TCAGGGAGGC CGTCACGTAT TTCCTATGGA ATTGGCTCCC TTGCAAAAGG CCTTCCGATT 7500

TGCTGGTTTT GAGCAGGTCA AGGCTGGAAT TGTCAAGTGG CCCATGTCTG TCCTACGTTT 7560

GACTTCGGAT TCCAAAGAGG ATTTGATCAA TTTGGCTGAT AAGATTTTGC AGGAATGGCG 7620

CCAGTATTCA GATCCTGCAG TGCAGATTTT GGCAGAGACA GACAGGACAC CGCATCACAC 7680

TATCACACCC ATTGCCCGCA AACGCGATGG ACAGTTTGAG TTGGACTTGG TCTTGCGAGA 7740

CAATCAGACT TCAGCAGAGT ATCCTGATGG TATCTATCAT CCCCACAAGG ATGTCCAACA 7800

TATCAAGAAG GAAAATATCG GCTTGATTGA GGTCATGGGC TTGGCAATCT TGCCACCACG 7860

TCTGAAAGAA GAAGTGGAGC AAGTCGCTAG CTATCTTGTA GGAGAAGCTG TTACAGTTGC 7920

CGATTATCAT CAGGAGTGGG CAGACCAACT CAAATCCCAA CATCCAGACT AACGGATAAA 7980

GAAAAAGCCC TTGCAATCGT CAAGGACTCT GTGGGTGCTA TCTTTGCGCG TGTACTTGAG 8040

GATGCAGGAG TCTACAAGCA GACAGAACAA GGGCAGACAG CCTTTATGCG CTTTGTGGAA 8100

CAGGTCGGAA TTTTACTAGA CTAGGAGCTT TCTCGG 8136

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10011 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

CCCATAGTGA AGAGTGGCCA TAAGAAGGTC TTCTAGGCTT AATTTAGGTT TTCGTCCACC 60

TTTTGCGTGT TTAAGTTGAT AAGCTGTTTT TAACACAGCT GAACATCTCT TCAAAAGTCG 120

TGCGCTGAAC ACCAACAAGA CATTTAAATC GTGTATCAGT TAGTTGTTTA CTTGCTTCAT 180

CATTCATAGA ACTACTATAC CATGTTTTGT TTCGCAGGAA GTCTAATATT GTCAAATACT 240

GGAACGCTCA TTGCTGGGAT ACGGAATAAG ATTGGCCCAG CTTCGATAAC TGGGATACCT 300

GGTTCAAAAC CAAGGTCTGT TGCAGCGATT GGTGTAAAGA TATCGTAACC TTTCATAAGG 360

TCTTCGTTTA CATCTTTCAC CATAACTGCA TCACAGTGAA CATCGTAACC ACGGTTTGAA 420

AGTTCTTCTT CTAGAGCACT TTTAATTTGG TGACTTGAGT TAACACCTGC ACCGCAGGCA 480

GCAAGAATTT TAATCATTTG GATTTCCTCC GATTTTATTT TTTAATAGAC AAGATTAAGC 540

GGTTGCTTCA GCAATGTAAG CATAAAGGGC TTCTGGTTCA GAAATTTTTG ATAGGTCTTC 600

AAGATGACCA TTTCCTGTGA AGAAGTCCAT TAACTGAGCA AGAATGTTCG TTTGACTTGA 660

ACTTGAATTA TTGATGATAA AGAAGAGCAA GGATACTTCT ACTTCCTTAC CTGGCGCAAT 720

CATATTATGG AAAGTCACCG GTTTCTCTAA TCGAACAACC ACCACTTTCT CAGCTAGATT 780

ATGAACAATA TCTGTGTGAG GAATCATTAC ATTTGCAAGT CCTTTCCTAG AAATTCCATA 840

TATAAACCAG TTGGAAATGA CTTTTCACGC GTGATCAAGG CTTCACGATA AGTTGGAGTG 900

ACAATTTCTC GTTCTTCCAA CAAGCTTGCT ACCTGATCAA AAAGTTATTC TTGATTATCC 960

GCTTCTAAGC AAAACACAAG GTTTTTGTCA AAGAAATAAT CTAATACCAT AAGGTTTTCC 1020

CTTCTTTCCA TTAACTTTAT GCTATAAGTA TAACACTATA TGAAATCGTT GTTAATTACT 1080

TTCTATTCTT TTTTGTCTCT TTTTTTATAT TTTTGTTTTG TTTATAGTTT GTTATATAAA 1140

AATAAACACA CAAACAAATA CTCCAAGCAT TTTTCTGTTC TAATACTCAA TGAAAATCAA 1200

AGAGCAAACT AGGAAGCTAG CCGCAGTTGT TCAAAACACA GTTTTGAGGT TGTAGATGAA 1260

ACTGACGAAG TCACTCAAAA CATGGTTTTG AGGTTGTAGA TGAAACTGAC GAAGCAACAg 1320

CCATACATAC GGTAAGGCGA CGCTGACGTG GTTTGAAGAG ATTTTCGAAG AGTATAAAAA 1380

CTAAAAAAGC AGACCATCTA AGCCTGCTTT ACTATTGATT CTTATATAAA TTTCCTGTGA 1440

ACAAGGAAAG GCATTTCTGA TAACTTATTC TTCATCCATA CTCAAGACGC TGAGGAAGGC 1500

TTCTTGCGGA ACTTCAACTG ATCCGATGGA TTTCATGCGT TTCTTACCAG CTTTTTGTTT 1560

TTCAAGGAGT TTACGCTTAC GAGAAACGTC ACCACCATAA CATTTAGCAA GTACGTTCTT 1620

ACGAAGGGCC TTGATATCAG TACGAGCGAC AATCTTGTGT CCAATAGCCG CTTGGATTGG 1680

AACTTCAAAT TGTTGGCGAG GGATGATTTT CTTGAGTTTA TCAACGATGA GTTTCCCACG 1740

TTCGTAGGCA AAGTCCTTGT GAACGATAAA GCTGAGGGCA TCCACCTTAT CTCCATTGAG 1800

AAGAATATCC ATTTTCACCA GCTTAGATGG GCGATATTCT GACAATTCGT AGTCAAAGCT 1860

TGCATAACCA CGTGTCGAAG ACTTAAGTTT ATCAAAGAAG TCAAAGACAA TTTCAGCAAG 1920

AGGAATTTGA TAGATAACAT TGACACGGTT ATCATCAATA TAGTCCATAG TCACAAAGTC 1980

CCCACGCTTA CGCTGAGCTA GCTCCATTAC TGCTCCGACG AACTCCTGTG GTACCATGAT 2040

TTGCGCCTTG ACATAAGGCT CTTCAATGGT CGCAATCTTA GTTGGGTCTG GAAACTCAGA 2100

TGGGTTAGAC ACATCCATAG ACTCACCGTC GGTCAAATTA ACTTTGTAAA TAACAGACGG 2160

AGCTGTCATG ATGAGGTCAA TATTGAACTC ACGCTCTAAA CGTTCCTGGA TAACATCCAT 2220

ATGGAGAAGT CCAAGAAATC CACAACGGAA ACCAAATCCA AGTGCCTGAG ATGTTTCTGG 2280

TTCAAACTGA AGACTAGCAT CATTCAGTTG CAATTTTTCA AGcGCTTCAC GCAGGTCATT 2340

GTACTTGTTT GATTCGATTG GGTAGAGACC CGCAAAGACC ATAGGATTCA TCTGCTTATA 2400

ACCATGTAAT GGTTCTGCCG CAGGATTGGT TGCCAAGGTA ACGGTATCAC CCACACGAGT 2460

ATCCTGAACC GTCTTGATAG ACGCCGCAAT GTAACCAACA TCACCAGTCG CAAGGAAATC 2520

ACGACCAACC GCTTTTGGTG TAAAAATACC GACTTCGGCC ACATCAAAGG TCTTACTATT 2580

GCTCATGAGC TGAATCTTAT CACCAGGTTT GACCACTCCG TCCATGACAC GCACTTGGAG 2640

GATAACCCCA CGGTAAGCAT CGTAAACAGA GTCGAAAATC AAGGCCTTAA GTGGCGCCGT 2700

CACATCACCC GTTGGTGCTG GTACTTTTTC TACAATTTGC TCGAGGATTT CTTCAATCCC 2760

AATACCAGCC TTGGCAGAAG CCAAAACTGC TTCACTGGCA TCCAAACCAA TCACATCTTC 2820

AATCTCTGTA CGCACGCGCT CCGGATCTGC AGCCGGCAGG TCAATTTTAT TAATGATAGG 2880

CATGATTTCC AAATCATTAT CCAAAGCCAG ATAAACGTTG GCAAGAGTTT GAGCCTCAAT 2940

TCCTTGAGCC GCATCGACCA CCAAAATAGC ACCCTCACAG GCAGCTAGCG AACGTGAAAC 3000

TTCATAGGTA AAGTCAACGT GCCCTGGTGT GTCAATCAAG TGGAAAATAT AAGTTTCCCC 3060

ATCTTTTGCA GTGTAATTCA ACTCGATGGC ATTCAACTTA ATAGTAATTC CACGTTCCCG 3120

CTCTAGCTCC ATGCTATCCA AAAGCTGGGC CTGCATTTCA CGACTTGAAA CCGTCTCTGT 3180

TTTTTCCAAA ATGCGGTCTG CTAGAGTTGA TTTTCCGTGG TCAATATGGG CGATAATAGA 3240

GAAGTTACGG ATCTTCTCCT GTCGTTTTTT CAATTCTTCT AAGTTCATGA TTCTCTTCCT 3300

TTCAGGGTAT CTATTTATTA TAAATTGTTT TTGATATTTT GACAAGACCA TACCCTGCTA 3360

GGAGTACTAA TCTTCAGCGA CAAAGCCGTC ATTTTCGATA AAGTGGTGTT CTGTCATTCC 3420

TTGGTCTGTA AAGACAATCC CGTGAAGGAC ACCACCATAA ACAGCTCCTC CATCCATTCC 3480

AATCTTGCCA TCTTCTGTAG TCCAAAGCTC AGATGTACCG CGTTCTTGCT GTAACAAACC 3540

ATAGACCGGT GTATGACCGA AGACAATGGT TTTTCCAGTA TGATTTTCAG CTCCGTGGAA 3600

TGGTTTTCTA AGCCATACTT TTTTATAATC TGTTGTTTCA TGCCAGTCGT CCAAGGTCAA 3660

ATCAATACCT GCGTGAACAA AGATATACTT GTCTGTCTCT ACTACAAATG GCATTTGACG 3720

AATGAATTCG ACCAAGTCTG CCGCTTCAGC GgCAACCCGC TTGGCATCTT CTACTCCATC 3780

AACTGGTGCA TCCAAGGGAC GACCTAGGAT AGAGTTAATG GTTGTATCTC CACCATTGCG 3840

ACTATAATGG TCATAACTTT CTTCTGGGTC ATCTAGCCAA GTCAAAAACA TATACTCGTG 3900

GTTTCCGGAC AAACAGATAG CCCCTTGATT GTCCACCAAG TCCTTGACCA TTTCAAGAAC 3960

ACGGTGACTA TCCTCACCTC TGTCAATCAA ATCACCTAGA AAGAGCAACT GGGGCTGACC 4020

ATCCCAGGTT TTGAGAAGGT CTTCCAGCAT CCCAGCTTTT CCGTGAACAT CTCCAATTAC 4080

ATAATAATCT GTCATCTTAT TTCTCCCTGT TTCTCAACAA TTCTCTTGCT TGCGTCAGGG 4140

CTGCTTCTGT CACATCATCA CCTGCCAACA TCTTGGCAAC TTCCTCCACT CGCTCTTCGA 4200

CCGTCAAGAG ACGAACAGTC GAAACCGTTG AATGGTCATT ACTAATCTTC TCAATAAAGA 4260

ATTGATAATC TGCAATCGCA ATTACTTGTG GCAAATGGGA GATAGCCAAA ACCTGACCAT 4320

GCTGACCAAT TTTATGAATT TTCTGAGCAA TAGCTTGAGC AACACGACCT GAAACTCCCG 4380

TATCCACCTC ATCAAAGACA ATGCTAGTCT TGCCTTCTTT ACGTGAAAAG GCAGACTTAA 4440

TGGCTAACAT GAGACGAGAT AATTCCCCTC CAGAAGCAAC CTTAACCAAG GGTTTAAAGT 4500

CTTCTCCAGG GTTGGTTGAA ATATAAAACT CAACCATTTT ATTTCCCTCA CGACTGAATT 4560

TTCCCTTACT AAAACGAACC TGAAACTGGG CTTTTTCCAT ATAAAGATCT TGCAGTTCTT 4620

GTTTAATCTC AGCTTCGAGT TGCTGAGCCA AATTATGACG AGCAGAAGCA AGTTGACCTG 4680

CCAAATTGAC AAGATTGACT TCCAACTTCT TAAGCTCTGC TTCCATGTCC TCAGACGAAA 4740

GATTATTGCC TGTCAAGAGA TTGTATTCTT CCGTAATCTT GGCAAAATAA AGCAAAACAT 4800

CATCAACAGT CCCACCATAC TTACGAGTAA TAGTATGAAG GAGGTCCAAA CGATTCTCAA 4860

CCTGCATCAG GCGATTGCCA TCAAAATCAA GGTCCTCAAT GATAGCTTCC AAACGTTTGC 4920

TAATGTCTTC TAAAACATAG TAGGTCTCAG ACAGATAGCT TGAAATTTCA CGGTATTCAG 4980

GATCATACTC TTCGACACTT TCCATGTCAT TCATAGCTGA ACGAACATTG GCCAGACTTG 5040

AAAAATCTTC ATTGTCCAAC ATACTGTAGG CATTGGTCAG TGTATCCGCA ATATTTTTGT 5100

GGTTGAGGAG TTTATCTCGC TCTTGATTGA GAGCCAAGTC TTCTCCAGCC TGCAAGTTTG 5160

CTGCCTCAAT CTCTGCCATT TGAAATTCCA ACATTTCGAT ACGTGCCTTG TGTTCCTGTT 5220

GGTTTTTCTT GACTTCCAGA ACCTGCTTGC GCATTTTCCG ATAGGCATCA AAACTCGTTT 5280

GATAGGTTTC TTTCAAGTCC CAAAAAGCGG CATCACCAAA TTCATCCAAC ATCTGGATAT 5340

GCAGTTGGGG ACGCATTAAC TCCTCATGGT CATGCTGACC ATGAATATCT ACAAGATGTT 5400

GCCCAATAGC TCGCAAAACA GACAGATTAA CCATCTGACC ATTTACACGG CTGATACTAC 5460

GACCATTTTG CAAGATTTCC CGACGGATGA TAATTTCATC ACCTAATTCT AAACCTTGCT 5520

CATCAAAAAT TTCCTGTAAA AGACGACTAT TCTCAACTGA GAAAAGCCCC TCAATCTCTG 5580

CCTTTGGTGC ACCATGACGA ATAACATCTG TCGTCGCACG AGCTCCCAAC ATCATATTCA 5640

TGGCATCAAT GATAATCGAC TTCCCTGCAC CCGTTTCACC AGTCAGGACA GTCATCCCCT 5700

TTTCAAAATT GAGGGAAATA GCCTCAATAA TGGCAAAGTT TTTTATCGAA ATTTCAAGTA 5760

ACATATAGAC CTACCAATTT TTTACTTGTT CAAAGATTTC CTCTGCTAGA CTTCCACTTC 5820

TGGCAATGAC TAAAATCGAG CTATCATCAG TCAAACAGCT AAAAATCTTG TCTGCAAAAG 5880

TCTCGATTAA CTGAGCTTTT ACAAAAGCCG TATTTCCTGG AATAACTTGG AGATTGATCA 5940

TCTTATCCAT CAATTCAGCC GATTCGATAT TGTCTTCAGC CAGTTGCAGA CTTTTTACGA 6000

TTGATTTTGG CAATTCGTAG ACATAGGTGT TGTCTCTCAA AGGAATTTTG ACAATACCTA 6060

ACTCTTTGAT ATCTCGGGAT ACCGTCGCCT GAGTGGCAGT GATACCTGCT TCTTTCAAAT 6120

GTTCTACAAT TTCTTCTTGC GTGCCGATTT GATAATCTGT CACCAATCTT CTAATTTTTT 6180

CAAGTCTCTC TTTTTTATTC ATTTTTAAAT TGACTATGCG CCCTCTCTAC TGCTTCTTTA 6240

ATCTCAGCAA GAATCTGATT GCTTGCTGAC TTTTCTTTTT TCAAATACGC TAAAAATTCA 6300

ATATTTCCAT GTCCACCTTG GATGGGAGAA AAGTCCAAGC CAAGGACTGA AAAACCTACC 6360

TCTACTGCCA TAGCTGTTAC AGATTCAAGG ACATTCTGAT GAACCTTAGC ATCTCGAATA 6420

ATTCCATTTT TCCCAATCTG CTCACGTCCT GCCTCAAACT GAGGTTTGAC AAGTGCTACC 6480

ACCTGACCTT GATCAGCCAA GACACGGTGC AAGGCTGGCA AAATCAGACT AAGGGAAATG 6540

AAACTCACAT CAATACTGGC AAAGCTCGGC TCCTGCTCGA AATCAGTCTT TTCAGCATAG 6600

CGGAAATTGA ACTGCTCCAT GCTGACAACT CGTGGGTCTT GGCGTAATTT CCAAGCCAAC 6660

TGATTGGTAC CAACATCGAC TGCAAAGACC AACTTGGCAC TATTCTGTAG CATGACATCG 6720

GTAAAACCTC CAGTAGAGGC CCCGATATCA ATCGTAGTCG CGCCATCCAC CGACAAATCA 6780

AAGACCTGCA AGGCCTTTTC CAGTTTCAAA CCACCACGGC TGACATACTT GAGTTTCTCC 6840

CCCTTGAGTT TTAATTCGGT GTCATCTGGA ATTTTCTCTC CTGGCTTGTC AAACCGTTCT 6900

CCATTAAGGA CTGCTACGAC TAGGCCAGCC ATCACACCTC GCTTGGCCTG CTCTCTCGTT 6960

TCAAACAACC CCTGTTTATA AGCTAGTACA TCCACTCTTT CCTTAGCCAT TGATTCTCAA 7020

ACTTTCTACT ACACTTACAA TCGATTCTGT TTCAAAGGGA AGCTGCTGGG CAATTTCTTC 7080

TAATTTTTCA TTAGCTTGAT CCAGGGTTTG GTTACAAAAG GCAATGGACT CTTCCAAGCC 7140

CAACAGGGCA GGATAGGTTG ATTTTTCTGC CTGCAGATCC TTTTGAGGTG TCTTGCCGAT 7200

TTCCTCAAAA CTAGCTGTCA CATCCAGTAC ATCATCTCTG ACTTGAAAAG CAAGTCCAAT 7260

CAATTCACCC ACAGTTTTCA GCTTCACCTG CATTTCAGGT GACAATTCAG CTATAATAGC 7320

TGCCGCTTGG AAGGGATAGG CTAGTAACTT CCCAGTCTTA TTGGCATGAA TAGTCTGAAG 7380

TTCTTCCAAA GACAAGTGCT GGTGTTCGCC CTCCATATCC AAAACTTGCC CTGCTACCAT 7440

ACCCAGACTA CCTGAAGCAA GGGATAAGTT GGCAATCAAG TCCACCTTAA TCTGACTTGG 7500

CAAATCTGCC TGCGCAATCA AGGCATATGA GTCTAAGAAT AAGGCATCTC CAGCCAAAAT 7560

GGCCATAGCT TCACCGAATT TCTTGTGATT GGTTAACCGC CCTCTTCGAT AATCGTCATC 7620

ATCCATAGCA GGAAGGTCAT CGTGAATCAA GCTCCCTGTA TGAATCATCT CTAAGGCAGT 7680

AGCTACCTGC GCGTGAGCAG GTTTGATGGT AACCTGCAAG GCTTCCAGAA CTTCTAACAA 7740

GAGAAAAGGC CGAATACGCT TGCCACCAGC ATGAATAGAA TAGAGAACAG ACTCCCGTAA 7800

ACTAGAGGCA AACTGCTGGT CTCCATAAAA ATCTTCCAAA GCCGACTCGA CAAGAGCTAA 7860

TTTTTCTTGC TTTTTCATTC AAAATCACTT TCTGTTCCGT CTTCTTGCAT GACCTTGACC 7920

AAGGTCTTTT CAGCCTTGTC CAGCGTAGCT TGGAGCTCTT TTGACAAGAC CATGCCCTTT 7980

TGAAAGGCAG TAATCGCATC TTCCAGAGCA ATTTCACCAT TTTCCAAACT TTGGACAATG 8040

GTTTCCAGTT CTGCTAGATT TTCCTCAAAT TTCTTTTGTT TTGACATCTT TAACCTCTAA 8100

TTCTACTTGA CCATCTCGCA TCAAAAGCGT TACTTGGTCT TTTTTCTTCA AACTCTCAAC 8160

CGAATCTACA ACGGACTCTT CTTTTTTGAC AATAGCATAA CCACGCGCCA CGATTCGGCT 8220

AGTATCCAAC ATGAGCAAAG CTTCCGAAAG TCGCTTGGCC TCAGCAACCT TGGCGTCATA 8280

AACTAACGCC ATTTGGCTAC CTAAGAGCTT GTCCAACTGT CCTAAACGGT CTTGATAGCG 8340

TTGGATTTTG GTAACAGGTG ATAATTGTAC TAATTGATGA GTTCTTGCTT GAACTAATTG 8400

TTTGTTATCA GAAATCCGAG TTCGCAAACT TTGTTTCAAA CGCAGTTGCA GTTGGTCCAA 8460

GCGTTGCAAA TAACCGTCAT ACAAGCGCTC AGGTTGTCTA AAGATAACAG ACTGACTGCA 8520

TTTTTTCAAA GCCTCTTGTT TCTTAGATAG AACATTTCGG ACTGCCGTTA CCATCCGTTT 8580

TTCCTGATTT TGCAAATGAG CTAATACATC CAACTTGGTC ACAGGTGTTG CCAGTTCAGC 8640

CGCCGCTGTT GGCGTTGCAG CGCGTCGATC TGCCACAAAA TCTGCCAAGG TCACATCCGT 8700

CTCATGCCCC ACACTAGAGA TAACTGGCAA ACGAGATTCA AAAATAGCTC GTACCACAAT 8760

TTCTTCGTTA AAGGCCCAGA GATCCTCAAT AGAACCACCT CCACGACCAA TAATGAGCAA 8820

ATCCAAATCG TCCCGTTGAT TAGCACGCGC AATATTTCTA GCAATTTCCT CCGCAGCCCC 8880

TTCACCTTGA ACCTTGGTCG GATAAAGAAG GATGTCAACA CCTGGGAATC GCCTGCTGAC 8940

GGTCGTGATA ATATCTCGAA TAACGGCTCC ACTACGGCTG GTTACTACAC CAATTCTCTT 9000

AGAAAATTGG GGCAGAGCTT GCTTGAAGCG TTCTTGAAAC AGGCCTTCTT CTGTCAATTT 9060

TTTCTTAAGT TGTTCAAACT GAATCGCAAG CGCCCCAACC CCATCAGGCT CAGCTTTTTC 9120

AATGATGATG GAGTAGCTAC CACTTGGTTC ATAGACCTGT ACACGCCCAA TCACATTGAT 9180

CTTCATTCCT TCTTCCAGGT CAAACCCTAA TTTCTGATAA ATCCCAGACC AGATGGTCGC 9240

TTGAATAACT GCATGGTCAT CCTTTAGGGA GAAATATTGG TGAGTAGGTC GTTTACGAAA 9300

GTTGGAAACT TGACCAGTTA AATAGACCCG TTCCAAGTAT GGGTCTTTAT CGAATTTCAT 9360

TTTCAGATAC TTGGTCAAAG TTGTTACCGA TAAATACTTT TCCATCTCCA CCTACTATTC 9420

ATTTACTTGC TCTTTCATGG GTATTATTAT ACCAAAAATA TGCCTAAAAA TCTCCATTTA 9480

TGTACCATTA TGAGGGAAAA ATAGAAAAAG GAGGCAAGGC CTCCACATGT GATTATTTGC 9540

TGTTTCGAGC TTCTTCCAAA ATCTTTGCAA TCTTGGTCGT CAACAGGTCG ATAGCCACGG 9600

TATTGCTAAC CCCTTCAGGA ATGACGATAT CAGCATAACG CTTAGTTGAC TCGATAAACT 9660

GGTGGTACAT TGGTTTGACC ACACCTAAGT ACTGGTTAAT AACGCTATCA AGGCTACGGC 9720

CACGCTCCTC CATATCACGC TTGATACGAC GAATAATGCG CACATCGTCA TCCGTATCCA 9780

CAAAAATCTT GATATCCATC AAATCGCGCA GACGCTTGTC CTCCAAGACC AAAATACCCT 9840

CAACGATAAA GACATCTTGA GGTTCCTGAC GATAGGTCTT GCTACTCCGT GTATGCTCTG 9900

TATAGTCGTA GGTCGGGATG TCCACCGGAC GCCCTGCCAA CAATTCCTTA ATCTGCTCGA 9960

TCATCAAGTC TGTATCAAAG GCAAAAGGAT GGTCATAGTT GGTTTTGACG G 10011

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5365 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

CGTGTGGTCT TAAAAATAGA AGACAAAGAA CAAACTGTTG GAGGCTTTGT CCTTGCAGGC 60

TCAGCCCAAG AAAAAACCAA AACAGCTCAA GTTGTGGCTA CTGGACAAGG TGTTCGTACC 120

TTGAACGGTG ACTTGGTTGC TCCAAGTGTT AAAACTGGAG ATCGTGTCTT AGTTGAAGCC 180

CACGCAGGTC TTGATGTCAA AGATGGCGAT GAAAAGTACA TCATCGTAGG CGACTAACAT 240

TTTGGCAATC ATTGAGGAAT AGAAGGAGAA AGTAAGTATG TCAAAAGAAA TTAAATTTTC 300

ATCAGATGCC CGTTCAGCCA TGGTTCGTGG TGTCGATATC CTTGCAGACA CTGTTAAAGT 360

AACCTTGGGA CCAAAAGGTC GCAATGTCGT TCTTGAAAAG TCATTCGGTT CACCCTTGAT 420

TACCAATGAC GGTGTGACCA TTGCCAAAGA AATCGAATTG GAAGACCATT TTGAAAATAT 480

GGGTGCTAAG TTAGTATCAG AAGTAGCTTC TAAAACCAAT GATATCGCAG GTGACGGAAC 540

TACGACTGCA ACAGTCTTGA CCCAAGCTAT CGTCCGTGAA GGAATCAAAA ACGTCACAGC 600

AGGTGCAAAT CCAATCGGTA TTCGTCGTGG GATTGAAACA GCAGTTGCCG CAGCAGTTGA 660

AGCTTTGAAA AACAACGCCA TCCCTGTTGC CAATAAAGAA GCTATCGCTC AAGTTGCAGC 720

CGTATCTTCT CGTTCTGAAA AAGTTGGTGA GTACATCTCT GAAGCAATGG AAAAAGTTGG 780

CAAAGACGGT GTCATCACCA TCGAAGAGTC ACGTGGTATG GAAACAGAGC TTGAAGTCGT 840

AGAAGGAATG CAGTTTGACC GTGGTTACCT TTCACAGTAC ATGGTGACAG ATAGCGAAAA 900

AATGGTGGCT GACCTTGAAA ATCCGTACAT TTTGATTACA GACAAGAAAA TTTCCAATAT 960

CCAAGAAATC TTGCCACTTT TGGAAAGCAT TCTCCAAAGC AATCGTCCAC TCTTGATTAT 1020

TGCGGATGAT GTGGATGGCG AGGCTCTTCC AACTCTTGTT TTGAACAAGA TTCGTGGAAC 1080

CTTCAACGTA GTAGCAGTCA AGGCACCTGG TTTTGGTGAC CGTCGCAAAG CCATGCTTGA 1140

AGATATCGCC ATCTTAACAG GCGGAACAGT TATCACAGAA GACCTTGGTC TTGAGTTGAA 1200

AGATGCGACA ATTGAAGCTC TTGGTCAAGC AGCGAGAGTG ACCGTGGACA AAGATAGCAC 1260

GGTTATTGTA GAAGGTGCAG GAAATCCTGA AGCGATTTCT CACCGTGTTG CGGTTATCAA 1320

GTCTCAAATC GAAACTACAA CTTCTGAATT TGACCGTGAA AAATTGCAAG AACGCTTGGC 1380

CAAATTGTCA GGTGGTGTAG CGGTTATTAA GGTTGGAGCC GCAACTGAAA CTGAGTTGAA 1440

AGAAATGAAA CTCCGCATTG AAGATGCCCT CAACGCTACT CGTGCAGCTG TTGAAGAAGG 1500

TATTGTTGCA GGTGGTGGAA CAGCTCTTGC CAATGTGATT CCAGCTGTTG CTACCTTGGA 1560

ATTGACAGGA GATGAAGCAA CAGGACGTAA TATTGTTCTC CGTGCTTTGG AAGAACCCGT 1620

TCGTCAAATT GCTCACAATG CAGGATTTGA AGGATCTATC GTTATCGATC GTTTGAAAAA 1680

TGCTGAGCTT GGTATAGGAT TTAACGCAGC AACTGGCGAG TGGGTTAACA TGATTGATCA 1740

AGGTATCATT GATCCAGTTA AAGTGAGTCG TTCAGCCCTA CAAAATGCAG CATCTGTAGC 1800

CAGCTTGATT TTGACAACAG AAGCAGTCGT AGCCAATAAA CCAGAACCAG TAGCCCCAGC 1860

TCCAGCAATG GATCCAAGCA TGATGGGCGG GATGATGTAA GCTTTCTATA GAAAACAACT 1920

TATAAAAAAC ACAAAAGGAG GGAATGACTA ACCCTTCTTT TTATAGGCTC TTTGTCAACT 1980

GTAGTGGGTT GAAGTCAGCT AAGCTCGAGA AAGGACAAAT TTCGTCCTTT CTTTTTTGAT 2040

GTTCAAAGCG ATAAAAATCC GTTTTTTGAA GTTTTCAAAG TTTCGAAAAC CAAAGGCATT 2100

GCGCTTGATA AGTTTGATGA GATTATTGGT CGCTTCCGGT TTGGCGTTAG AATAGTGTAG 2160

TTGAAGGGCG TTGATAATCT TTTCTTTATC TTTGAGGAAG GTTTTAAAGA CAGTCTGAAA 2220

AATAGGATGA ACTTGCTTAA GATTGTCCTC AATAAGTCCG AAAAATTTCT CCGGTTCCTT 2280

ATTCTGAAAG TGAAACAGCA AGAGTTGATA GAGCTGATAG TGATGTTTCA AGTCTTGTGA 2340

ATAGCTCAAA AGCTTGTCTA AAATCTCTTT ATTGGTTAAA TGCATACGAA AAGTAGGACG 2400

ATAAAATCGC TTATCACTCA GTTTACGGCT ATCCTGTTGT ATGAGCTTCC AGTAGCGCTT 2460

GATAGCCTTG TATTCATGGG ATTTTCGATC CAATTGGTTC ATAATTTGAA CACGCACACG 2520

ACTCATAGCA CGGCTAAGAT GTTGTACAAT GTGAAAGCGA TCCAACACGA TTTTAGCATT 2580

CGGGAGTGAA ACAGTCTGGG AGACTGTTTC AGCCTGAGCC TAGAAATTTG AAAGCGAAGC 2640

TGTTTAGCCA AGTCATAGTA AGGACTAAAC ATATCCATCG TAATGATTTT CACTTGACAA 2700

CGAACGGCTC TATCGTAGCG AAGAAAGTGA TTTCGGATGA CAGCTTGTGT TCTGCCTTCA 2760

AGAACAGTGA TAATATTAAG ATTATCAAAA TCTTGCGCAA TGAAACTCAT CTTTCCCTTA 2820

GTGAAGGCAT ACTCATCCCA AGACATAATC TTTGGAAGCC GAGAAAAATC ATGCTCAAAG 2880

TGAAAGTCAT TGAGCTTGCG AATGACAGTT GAAGTTGAAA TGGCCAGCTG ATGGGCAATA 2940

TCAGTCATAG AAATTTTTTC AATTAACTTT TGAGCAATCT TTTGGTTGAT GATACGAGGG 3000

ATTTGGTGAT TTTTCTTTAC CAGGGGAGTC TCAGCAACCA TCATTTTTGA ACAGTGATAG 3060

CACTTGAAAC GACGCTTTCT AAGGAGAATT CTAGAAGGCA TACCAGTCGT TTCAAGATAA 3120

GGAATTTTAG AAGGTTTTTG AAAGTCATAT TTCTTCAATT GGTTTCCGCA CTCAGGGCAA 3180

GATGGGGCGT CGTAGTCCAG TTTGGCGATG ATTTCCTTGT GTGTATCCTT ATTGATGATG 3240

TCTAAAATCT GGATATTAGG GTCTTTAATA TCGAGCAGTT TTGTGATAAA ATGTAATTGT 3300

TCCATATGAA TCTTTCTAAT GAGTTGTTTT GTCGCTTTTC ATTATAGGTC ATATGGGACT 3360

TTTTTTCTAC AACAAAATAG GCTCCATAAT ATCTATAAGG GATTTACCCA CTACAAATAT 3420

TATAGAGCCG AAAATTCACA TCTAATATAT GCAGACTACT TTGAAATGAA ATTAAAAAAA 3480

TTATTAAAGG ATGACACAAA AGTTTTTGAA AAATCTACAT TCAAATTTGT AGAAGGATAT 3540

AAAATATACC TGACAGAATC TAAAGAATCT GGAATTAAAC AAATGGACAA TGTCATAAAA 3600

TATTTTGAGT TTATTGAATC TAAAAGTATT GCTTTATATT TTCAAAAACG ATTAAATGAG 3660

CTGATAGATT AAATAGCATT TTCTCTGTTG AGATATTGTT TTTAAAATAT TGTACTAAAT 3720

GATTGATGCT ATGTGGAAAT ACAAAAAAAT GTTTTTGATA CGAAGTTGAC CTGTATTTTT 3780

TATACTAATC ATTTTCGTAT TTTTTGTATT AAACGATATA AGTTTGTTGT AAACTTACAA 3840

GGAATAAAGA CATTAAAAAA TAACAGTATA TCTATTTGTT TTATATATTT TACGAATTCT 3900

GCATAAATCT CTTTCTAGTA ATGTGTTGTA ACTCTGCTAT AATAGATTTA TTCCTTTTTG 3960

TGTTTACACA ATTTATTTTA TAGTACCAAA AAAGGTCAGG ATTTTGTTCC TGACCTTTGA 4020

CAACTTTACC GATTCTTTAG TTCTACATAG CGCTTGTACC AAATGTTTAC ATAGGCTTCT 4080

GAGAAAGGAC CACGTCCATT GTTAATCCAA TCAACAAGAA TTTTGACATG TTCTTTTAAA 4140

ATATAGTCCA AGTCATCAGA ATAATTCATT TTGCGTTTGT GACGCTCGTA CTCTTCAACG 4200

TCCAAGAGAC GTTTTTCCCC ATCTGTAAAA ATTTTAACAT CCAAATCGTA ATCAATATAC 4260

TTCAGTGCTT CTTCATCCAG ATAGTAGGGG CTAGCCATAT TGCAATAGTA AGAAGTTCCA 4320

TTATCACGAA TCATGGCAAT GATATTAAAC CAATATTTCT TGTGAAAGTA AACAATAGCC 4380

GGTTCTCGAG TGACCCAACG ACGACCATCA CTTTCGGTAA CAAGTGTATG ATCGTTGACA 4440

CCAATAATGG CGTTTTCTGT TGTTTTTAGT ACCATGGTGT CCCGCCAAGT TCGGTGGAGA 4500

CTCCCATCAT GCTTATAACT TTGAATTGTA ATAAAGTCGC CTTCTTTTGG AAGCTTCATA 4560

ACTAACCAAC TTTCTACAAT TTATAAGTTT ATCATTTACT ATTGTACCAT AAAATTACCC 4620

AAAATCTGTG AATTTCACTT GGAAATATTA AAGATATTCT CTAAGAGCGC TTGCTATATC 4680

CGAAAAATCG TAGCCCTTTC GTGCTAAAAC TTGAGTTAAA CGCTGCTTCA GTTCGTATCC 4740

TTCATACTTT CGGGCATACT TAGTATATTG CTTATCAAGT TCCTTGAAGA TGAGTTCCTG 4800

AGTCGTTTCT TCATCAACTT GACTATCCAA TTCGTCAAAG GCAATTTTAG CATCAAAATA 4860

AGAGAAGCCC TTGTTAGTCA AGTTCTGGAT AATCTTATCT TGCAGGGCAC GAGCTGGAAG 4920

TTTTCCCTCA TATTTTTTCA ATAGTTTATT GGCTACACGT TGAGCAACTT CCGAAAAATC 4980

AAAATCATTC AAGATTTCTT CTATAGTAGA TTTTGAAATT CCTTTTTGTG CTAATTTCTG 5040

AGTCAGTACA TAAGGTCCCT TGTCTCCTGA AAGTTGATTG GCATTGATGA TAGCATAAGC 5100

GTACTGGCTA TCATTAATCC ACTTCTCTTC TTTAAGATTA GCAATGACTT GAGAAACGAT 5160

GTTTTCATTA ATATCATATT TTTTCAGATA TTCTCTGACC TCTTTTTCAG TACGTGCTTT 5220

AAAGGATAAG TGGTAGAGGG CCAGATTCTT ACCATAAGAA AATTGAGCAA AGTCTTGAAT 5280

CTCTTTCAAT TCCTCTTCGC TTATCACCTT ATCTCTCGAT AACATAAAAC GAACAATTGT 5340

GTCTTCGGTG ATATAGCATT TGTCG 5365

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3636 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

TTTCCAGAAA GAAGTTGAGT AAAGTCTTTA TCAAAGAGAA TGACTTCCGT ATTGGAACTG 60

ACATTAGGTT TTATTTCTAC TTTACTAGCG TCCGCCCTAG CATTTTCTAA ATCTTTAATC 120

TCTTCTGTTG CCCTATTTAT AGCCAGCTGA ATAACTGCTT GAGGATTTTC ACTCAGTCCA 180

TGAAGCTTAT CGTCCACCGA AGTATAAAGA CTCGAATGCA TGACTTGTAA AATAATCAGA 240

GTCATTGTAG AAAAAATCAG GGTGAAGACA CCGAAGTTGC GGATAAAATA ACTAAAGTCA 300

TCCGCATACC ATGTTTTTTT AAGTTTACTG AACATCTTTT AAAAGATACC CAACACTACG 360

CAAAGTTTGC AAATTCTCTG CAAAAGTGGT TCCCTTTAAT TTCTTACGGA CTTTTGAAAC 420

ATAGACTTCG ACAACCGAAA TCGTTGTATC ACTATCAAAT CCCCATAGAC GGTCAAAAAT 480

CTGCGTCTTA GGCAAAATCA CATTTTGATT TTGAAGGAAA TAAACTAGTA AATCGAACTC 540

TTTCCCCAGC AATTCGACAG GAGTATCTTC AACTTTAACG GTATTGGTTG ATAAATTAAC 600

CACGATATTC CCATAAGTCA AGGTGTTTTC ATTAAACTTC CCTGAACGTT TGAGAAGGGC 660

CTGAATCCGC ATTTTAAGTT CTTCTAGGTA GAAAGGTTTG GTCAGATAAT CATCCGCTCC 720

CAGTTCAAAT CCATGTCCCT TGTCATCCAA ACTTTCCTTG GCAGTCATAA TCAGAACTGG 780

TGTCGTAATT CCCTTTTCAC GCAATTCTTT TAAGACTTGG AAACCATTTT TTTCTGGCAA 840

CATCAAATCC AGCAAAATCA AGTCATAGAC ACCACTCTCA GCTTCGTAGA GACCTTCTTC 900

TCCATCAAAT ACCTGCATAA CATCCGCAAA ATCGTCTAAA AAGTCAAATA CTGAATTTGA 960

CAGACCTAGG TCATCCTCAA CCAATAAGAT TTTTATCATG AGAAACTCCT CCTTATTAAA 1020

ACTATTATAC CAAATTTGCC TTAAAAAAAA CTCAACTCTC TGCATTTTAC ATGAGATAGC 1080

TGAGTTTTCT TTTTATTTTA GGCTTATTTA TGCATTTCCG TATTGAAGAA CAACTGCTTC 1140

GACTGCAGCT TTTTCACGGC TAATCAAGTC AACACGCGCT GCAATTTCCT TGATTCCCAT 1200

ACCGATGTTA CGGCTAAGAG CAAGGTCAGA AAGTTGCGGT TCAAAGAACT CCTTGTATTC 1260

CGCCAAGCGT TGCTGAGTCT TAAATACATG AGCAGGAAGG ATAACAAAGC TATCAAAGCT 1320

CATATCTCCT CCAAGGGCTG CCTTAATCCA AGCCCAGTTT TCACGCGCCC AAGACCAAGC 1380

TGTTTTCTGA GTTGCTTGAT GAGCTAGGAA TTGGTAATAC CAAGCAGACA AGTCCTGTGG 1440

TTTGACCACA AATTTGTCCT TCCAAGAAGT AATCAGGTTT TGGATATTAT CCGCATCTGT 1500

ACTGTATGCA AGAGCTGCTG CCAACTGGCG TTTAAAGACA GCATCTGTTG CGTGAGTATA 1560

AGTATCAAGA TAAAGTGCTA ACAAGTCTTT AGTCTCATGA TGTTTCATCT CATTAATCAG 1620

AACTTGTGAG CGAATAGCTG CTGGGAGTCC TGCAAGATTC TCCTTGTGTG TTGCGAAGAT 1680

TTGGCTAGCG ACTTGACTAG CTTCTGCATC ATTTGAGCGA ATCATCATCG AAACAGCCAG 1740

CTGACGAACC AATTCATCCT CATCTGATTC TCCGTCTTTA GCTTCAAAAC CAAGACGGTC 1800

ATAGTTATGA CGAGCCAATT TAGCAACCAG TCCTTTGAAG GCTGTTTCAG CATCCGTTCC 1860

TTCATCAATA AAGCGCTCAA GGGCTGAAAT CACTTGAGAA ACAGCTGAAA CCACCAGATA 1920

AGACTCTTCC TTAGCAAGTT TATCAAGAAC TGGAAGCAAG TCTGCATAAG AAATGTGCCC 1980

TGCCTCAGCC AACAAACGAC GTTCTTGAAC AATTTGCAGT TTGCTTGTGT TATCAAGTGT 2040

CTCTAGCTCA GCAAGAACAG CTGCTAACAA GTCTCCTTGA TAGTCGGTAA TATAGTGGGC 2100

AGTATTTTCA GTGTTGAGAC GAAGAGCTCC TTCATTTTCA GCAAGAAGAG CTGCGTAGCC 2160

AGGGATTTCG ATACTTTCAG TTTCGAGTGT ATCAGGCAAG CCTTTCCAGT TGCTATTGAG 2220

GGGCACCACC CAGAGACGGT TCTTGTCTTC GTTCTCACCG ATGAAGAATT GTTTTTGTGA 2280

AATCTTCAAG ACATCATTTT CAACTTTAAC AGTAAGAACT GGGTAACCAG GCTGTTCCAA 2340

CCAAGAATCC ATGAAGGCTG CGACATCACG TCCTGACGCT TGACCAAGGG CATCCCAAAG 2400

GTCACTACCA ATGGTGTTGC TGTATTGGTG TTTTTCAAAG TAGGCGTGCA AACCTTTAGC 2460

AAAATCAGCA TCTCCTAGCC AACGGCGAAG CATGTGCATG AGACGGCTTC CTTTGGCATA 2520

GACGATAGCG CCGTCAAAGA GTGTATTGAT TTCATCTGGA TGTTTAACTT CGACGTGGAC 2580

AGACTGAACG CCATCAGTAG CGTCACGTTC AAGAGCAAGA GGTACTCCAC CTGTTTGGAA 2640

ATCTTCAAAG ATATTCCAGC TTGGTTCGAT GGTATCCACA CAGACGTATT CCATCATATT 2700

AGCGAAACTT TCATTGAGCC AAAGGTCATC CCACCATTTC ATAGTCACGA GGTTCCCAAA 2760

CCATTGGTGA GCCAATTCAT GGGCCACAAC AAGGGCAACT TGTTGACGGC TAGCAAATGT 2820

AGAGTTCTCA TCGACAACCA AGTAAACTTC ACGGTAGGTC ACAAGACCCC AGTTTTCCAT 2880

AGCACCAGCT GAGAAGTCAG GAAGGGCGAT GTGGAGAGAT TGAGGAATTG GGTACTTAAC 2940

TCCATAGTAA TCTTCGTAAA ACTCGATAGA GCGAACAGCG ATATCCAGTG AGAAATCAAG 3000

ATTTGAAAGT GGATGTGCTT TGGTTGAGTA GACACCTACC AGGGTACCAT TTTTAGTTTT 3060

AGCGGTCACC CCTTGCAAAT CACCAGCAAC AAAGGCCAAC AAGTAAGAAG ACATGCGAGG 3120

TGTTGTCTCA AACTTCCAGA TACCTGTTTC CTTACGGTTT TCAACATCGA TTTCTGGCAT 3180

GTTTGACAAG GCCAATTCAC CTTCTGCTTG GTCAAAGCGA AGAGAGACGT CAAAAGTTGC 3240

TTTGGCTTCA GGCTCATCCA CACATGGGAA AGcTTCGCGC GCAAAATGGC TCTCGAACTG 3300

AGTAGACAAG ACCTCCTTCT TGACTCCATC AACTGTATAA TAAGAAGGGT AAATCCCTGT 3360

CATGTTGTCT GTAATTTTAC CAGAAAAGGC AAGAACCAAT TCAACTTGAC CAGCCTCAGC 3420

CAATTCGATA TGAAGGGCTT CATTGTCATG GTCAACTGTA AATGGACGAG CTTGACCTGC 3480

AACTTCTACA GAGGTGATTT CCAAATCTTT TTGGTGGAGG GAGATGCGGT CACTCTGTGC 3540

TTGACCAGTG ATGGTCACTT TCCCAGAAAA AGTCTTGGTC TCACGACTCA AATCTAAAAA 3600

TAAATCATAA TGTTCAGGAA CAAATTGCTT AATGGG 3636

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5065 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

ATAGCGTGTA ATAATCGATT TTAGAGGTAC CATAAGCCAC CTCCTACAAA TAGAAACCGA 60

TATAAATCAA TGCCTTCCAC CCTTAGACTT CCCTAGTTCC TGTCTCAAGC GAAACATTTC 120

TTTGAAACAG GAATAAGTTA ACCAATTCAT ACCAATAGCT AGCAGAATAA AAAGAAACCA 180

AATGCCCCAT AACTTGATAT CTGTCACATT TCTCAAGACG GTATTGAAAA ACAGAACTGA 240

AACAACTGTC CAAGCAAGGC TAAAAAGAGA ATAGAAGGGG ATGTAAAACC AGTAAAAATA 300

ATAAAAAATT GGAAAAAACT TACTATTTCT GTTGGCCTTT TCAATCCAGT TATCAAAATA 360

AAAGTACGGT GCTAAAAGTA AGAATTTAAA CAAATGTTCC ATCACCGACA TCCCCCCTTC 420

TTTTGATAGC GTTTTCTATT ATTTTATTAT ATCAAAAAAA TCCGGAACTG TCATTCCAGA 480

TTCTACTTTT TTATTTGCGT TTTCTTGCGA TGAGATGAAT CGGTGTTCCC TCAAAAACAA 540

AGGCCTTGCG GATTTGATTT TCCAAGAAAC GCAGGTAAGA AAAGTGCATG AGTTCTTCTT 600

CATTGACAAA GATGACAAAG GTTGGTGGTT TGGTTGCCAC TTGGGTCGCA TAGAAAATCT 660

TGAGACGTTT TCCTTTGTCT GTCGGTGTTG GGTTGATGGC AATGGCATCC ATGATGACAT 720

CGTTCAAGAC AGCTGATGGA ATACGTGTAT TTTGACTTTC GCTGATTTGC TTAATCATCT 780

CAGGAAGTTT GTGGAGACGT TGCTTGGTTA AAGCTGATAC AAAGATAATC GGTGCGTAAG 840

GCAGGTATTG GAACTGCTCA CGGATATCTT CTTCCCAGTT TTTCATAGTG TGGTTATCTT 900

TTTCAAGCGT ATCCCACTTG TTGACCACGA TAATCATCCC TTTACCAGCT TCATGGGCAA 960

ATCCTGCGAT ACGCTTGTCG TACTCACGAA TGCCTTCTTC CGCATTGATG ACCATCAAGA 1020

CCACATCTGA ACGGTCAATA GCACGCATGG CACGCATAAC AGAGTATTTC TCAGTATTTT 1080

CATAAACCTT ACCAGACTTA CGCATACCAG CCGTATCAAT CATGGTAAAC TCTTGACCAT 1140

CTGTATCTGT AAAGTGGGTA TCAATGGCAT CACGAGTTGT TCCAGCAACA GGACTAGCAA 1200

TAACACGGTC TTCTCCCAAG ATAGCATTGA TCAAGCTTGA TTTTCCAACG TTAGGACGAC 1260

CAATCAAGCT AAACTTAATG ACATCTGGAT TTTCTTCCTC ATATTCATTT GGAAGATTTT 1320

CTACGATCGC ATCTAGCACA TCCCCTGTAC CGATTCCATG GACAGATGAG ATAGGCAATG 1380

GTTCACCCAA ACCGAGAGCA TAGAAATCAT ATATATCATT TCTCATCTCA GGGTTGTCCA 1440

CCTTGTTGAC TGCGAGGATA ACTGGTTTGT GGGTCTTATA AAGCTTACGA GCTACGTATT 1500

CGTCTGCATC AGTAATTCCT TCCTTACCAG ACACGACAAA AACGATAACA TCTGCTTCTT 1560

CCATGGCAAT TTCTGCCTGG TGCTTGATTT GTTCCATGAA AGGAGCATCG ACATCATCAA 1620

TTCCTCCTGT ATCAATCATG CTAAAAGAAC GATTGAGCCA CTCACCCGTT GCATAAATAC 1680

GGTCACGTGT CACTCCTTCG ACATCTTCTA CAATGGAGAT TCGCTCACCA GCGATCCGAT 1740

TAAATAGGGT TGATTTCCCA ACATTGGGAC GTCCTACAAT GGCAATAGTT GGTAGGGCCA 1800

TAATTTCTCA CTTTCTACAA TAATTTCTTC TGTTCAAGAT TTTTTCTAGT TGAGCTTGGT 1860

TCAGCTTGAC CAAACTGTTC TGCTAGGCGC TGACTCCAGC TTGTGGTCGC ACGCGCCCCA 1920

GCATAGTCAG CCTGAACACG GTCATAAGCT TGGATTGCCT CAGTTGACTG TTCTTGGTAT 1980

TCTTCCTCAA AGACAACATT CTCTAGTGGC AGTCTCGGTT TCATATCATG ATGTTGATTT 2040

GGCACACCCA GTGCCATCCC AAAGACAGAA TAGGTGTAGT CAGGTAGGTT AAAGAGCTCT 2100

GCCACTTCTT CAGACTTGTA TCGAACCAAA CCGATAATCA CACCACCATA GCCCAAGCTT 2160

TCAGCTGCCA ACAAGGCGTT TTGTCCAGCA AGAGCTGCAT CGACCGAACT AATCAAGAGA 2220

CCTTCCACAC CTTGGGGTTG GAAGGTGTCG GTATGAAGTC GGGCTCCCTT TTCTGCTCGG 2280

TTCAAATCTC CGACAAAGAG AAGGAAAACA GCAGACTGGC GAATGGCTTC TTGAGGTACC 2340

AATTCATACA AGGCATCTTT CTTCTCTTGA CTTCGTACCA CAATCACAGA GTAGGATTGG 2400

AAATTCTTCC AAGATGATGC CATCTGGGCT GCTGTCAAAA TCTCATTTAA GTCTACTTGG 2460

GGAATTTCTT GCTCTTTAAA CCTGCGCACT GAAGTATGAG CCTTCATCAA TTTAATGGTT 2520

TCTGTCATCG ACGGTTTACT CCTTCTAAAC GAGTCTCCTC AGCCAAATAA CGGATGCGTT 2580

CCATGACCCG TCTGGCTTCC CAGGTTTCGT CATTTCCATG TTTCACTTTC GCAAAATGCT 2640

TCTCCAAATC TTCAAAGTTG AAGTTGGATG TGAAAAAGGT CGGTAAATTT TCCTGCATCC 2700

GATATTGGAG AATGACCTGC AGGATTTCGT CACGCACCCA AACGGTTGAT TGCTCGGCGC 2760

CAATATCATC TAAAATCAGG ACCTCAGACA GCTTAATCTC ATCCACCAAG GTCTTAACAT 2820

TGCCATCACT GATAGCATTT TTGACATCAA TGACAAAGCT AGGATAGTGG AGGAGAGTTG 2880

ATGAAACACC ACGTTTTTCT GATAAATCAT GAGCTAAGGC CGCCACCATG AAACTTTTAC 2940

CCACACCAAA GTCTCCATAT AAGTAAAGAC CTTTTCGAAT AGCTGGATAT TGCTCCACGA 3000

AGGCTAGTAG CTTTTCAAAA ACTGGTAAGC GCCCCAAATC ATCCAAGTCA ACTTGAGCCA 3060

AACTAGCTTT CTTGAGACTG GCTGGTAGAT TGATTAACTT GAGACGGTTC TTAATAGCCG 3120

CTTCTTTTTC AGCCGCGATT AGCTCAGGAG TTTCTTCATA TGAAACATCT GCATAACCAT 3180

GATTCTTAAC CAAAATCGGC TTGTAGCCTT TGGCAATATA ATCCGTATCC CCACGGAGAA 3240

ACTTGTCACG CTCGGTGATG TACTGATTAA ACTTGGAGAT ACTGCGATTT AATTCCTTTG 3300

GAGTTAAGGA TTCTTGCTGG ATAAAGGCCG CAACATCAGG GTCCTTCATG ATTTTCTGGA 3360

CCAAATCTTG ATAATAAAAA CGGCTGGGTT GACGTTTGAG TACGTCTCCG ACACTTTCCA 3420

TCTAATCTCC TCCTTTTTCT AATCGAGCTA ATAGTTCTTG CTTCTTACGT TCTAGTTCCA 3480

GACGAGTTTC CTCGCTGGTT TCATTCTTAT ATTCAGGATT ACTCCATTTA GGAACATTGG 3540

TTTTTTCTGG GGCAGTCTGA TTCTGTTTTT GTGTTTTTGC TTTCTGCCCT CGATCACGAA 3600

TTCGTAAAAC GGCCTCTTCT GCCGAATGAA TCTTTTGATA GGCATAGTCA TTGGCTACCT 3660

TCATGGCATA TTTCTCATTG ATATTTGCCG AATCCACCTT ATTAAAGGTC AATAAGAGAA 3720

TAATATTGAT GACTTCGTCC AGTAAGCCCA AGCCAGCCAT CTGTTGCAAG AGTTCTCTTT 3780

CTGTTTGGGT AATGGTTCCC TTGCGTGTTT GCTTGATTTC TGCTAAGAAC TGCAGGGCAG 3840

TTTTACTTTT AGCTTCTTTG ATAATGGTCG CTTCCTTAAG ACTAAAGTCA GAGGAAACTG 3900

GTTTTTGAGC AATTTTTTCA CGCATGCGTT TGGTTGAAAT AACCTGGGAA ACAGCTGTTG 3960

ACTTGGCCAA TTGATAGGTT TCAAACCAAG TCCATTTCTT CTCCTCGGCA ATAGCAAAGA 4020

GGTTTAAGAC ATCGGACTGC TCATCCGCAA AACGAAGTCC ATCTCGAGCC ATCAGCTGGC 4080

GAAAATGTTC CAAGTCAAAA TCATTGGCCA CTTTCTTCTT GAGACCAAGG TCTTCTTGAC 4140

TGCCTAGTTC TGCCAATTCT GGAAAGACTT GATTGAGTGA GACAGGTATT TCTTCACCAT 4200

CAGCACTTTC AACTTTCAAA TCCTCCACAG CTACATCGCC AATCTTTTTC TCTAAGAGTC 4260

TGCGATAAAC AGGATGCCCC AAGAAGTCTT GACTAGATAG AGGAGCATGG AGGGCTAGCT 4320

GATAAACATC ACCCTTTTGA TAGAGGGTCA AGAGATTAAA AGCAGATAAG ATTTTCAATG 4380

ATTTTATCAG TCTATCCATC CCAAAGTTGA GATGGTTGAG AATGCTTGAA AAAAGATATT 4440

CCTTTCTACC ATTATCCCAA AAACTGATTG TATAAAGATA AAGGCTCAGT GCCTCCTGAC 4500

CGATAATCGG GAGGTAGCAC TGTACCAGAG ATGAGGTATC TTGCGACACC CGATTATTCT 4560

TTAGATAAGA AAAACGGTCA ATTGGCTTCA TTTATCTTTC CTTTTTCTTT TTAGAGGACT 4620

GGGTGATTTG TTGGAGCAAG CTCTCTAACT CACTGACATC CTTAAAACTA CGATAGACAC 4680

TAGCAAAACG TACATAGGTA ATCTCGTCCA ATTCAGCCAA CTCCTCCATG ACGAGTGAAC 4740

CAATGTCCTC ACTTTGAATT TCATTTTCAT TTCGACCACG GAGTTTCTGT TCGATACGAT 4800

TGACTACCAT GTTGATTTCA TCACTTGACA CAGGACGTTT CTGGGCTGAG CGGATAATCC 4860

CATTAAAGAT TTTATCTCTG GAGAATTGTT CCCGTGTGCC ATCTTTTTTA ACAACCACTA 4920

AGGTTCTTTC TTCTACTCGT TCGTAGGTTG TAAAACGGTG TTGGCATTCG TCGCACTCAC 4980

GTCTTCTACG AATGGTGTTC CCTTCTTCTG CTTGGCGACT ATCGATAACA CTTGACTTGG 5040

TAGCCCCACA TTTTGGACAG GGTACC 5066

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9607 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

CACTTGAAGT ATTTGAAACA GCTATGGAAA ACATCATGCC TGTACTTGAA GTACGTGCAC 60

GTCGTGTTGG TGGTTCTAAC TACCAAGTCC CAGTTGAAGT TCGTCCAGAA CGTCGTACAA 120

CACTTGGACT TCGTTGGTTG GTAACAATCG CTCGTCTTCG TGGTGAACAC ACAATGCAAG 180

ACCGTCTTGC AAAAGAAATC TTGGATGCTG CTAACAACAC TGGTGCAGCA GTTAAGAAAC 240

GTGAAGATAC TCACCGTATG GCTGAAGCTA ACCCTGCATT CGCACACTTC CGTTGGTAAG 300

ATAGGATGCG AAAGCGTTAA GAAAGTCCCA GAGAAAATAG GGAATCGAAG CAGGTTGCGG 360

TTGCAACCAA TGAGATTCAT CTTTTTCTCC AGACTTTTAG CTTGAGCTCA ACTAAATCAT 420

GATGCTAGGA ACGGTAAGGA TGCAAGGTAA AAATAGGAAA CTGACGCAGT ATTCGACGAA 480

TACAAGGAGT TTTATCTTTT TCACGCAGCA TCCCGTTCCA GCTCACATCG GCTAACTAAC 540

TTTAGCCCGG GTTCAAATTA GCTAAATCGA TTAGTATTAG CTATAACTCA GCTTACCATC 600

TCGTAAGTTG AAACCAACAA TAGCATGAAA ACATTGAGAA CGGGTAGGTC CTGCCTATCC 660

GTTTTTATTA AAATCGTGTT ATAATAGAAT AGAAATCAAA AATAAATAGG AGAAACAAAC 720

CTCATGGCAC GCGAATTTTC ACTTGAAAAA ACTCGTAATA TCGGTATCAT GGCTCACGTC 780

GATGCCGGTA AAACAACAAC TACTGAGCGT ATTCTTTACT ACACTGGTAA AATCCACAAA 840

ATCGGTGAAA CTCACGAAGG TGCGTCACAA ATGGACTGGA TGGAGCAAGA GCAAGAACGT 900

GGTATCACGA TCACATCTGC TGCGACGACA GCTCAATGGA ACAACCACCG CGTAAACATC 960

ATCGACACAC CAGGACACGT GGACTTCACA ATCGAAGTAC AACGTTCTCT TCGTGTATTG 1020

GATGGTGCGG TTACCGTTCT TGACTCACAA TCAGGTGTTG AGCCTCAAAC TGAAACAGTT 1080

TGGCGTCAAG CAACTGAGTA CGGAGTTCCA CGTATCGTAT TTGCCAACAA AATGGACAAA 1140

ATCGGTGCTG ACTTCCTTTA CTCTGTAAGC ACACTTCACG ATCGTCTTCA AGCAAATGCA 1200

CACCCAATCC AATTGCCAAT CGGTTCTGAA GATGACTTCC GTGGTATCAT TGACTTGATC 1260

AAGATGAAAG CTGAAATCTA TACTAACGAC CTTGGTACGG ATATCCTTGA AGAAGACATC 1320

CCAGCTGAAT ACCTTGACCA AGCTCAAGAA TACCGTGAAA AATTGATTGA AGCAGTTGCT 1380

GAAACTGACG AAGAATTGAT GATGAAATAC CTCGAAGGTG AAGAAATCAC TAACGAAGAA 1440

TTGAAAGCTG GTATCCGTAA AGCGACTATC AACGTTGAAT TCTTCCCAGT ATTGTGTGGT 1500

TCAGCCTTCA AAAACAAAGG TGTTCAATTG ATGCTTGATG CGGTTATCGA CTACCTTCCA 1560

AGCCCACTTG ACATCCCAGC AATCAAAGGT ATTAACCCAG ATACAGACGC TGAAGAAATT 1620

CGTCCAGCAT CTGACGAAGA GCCATTTGCA GCTCTTGCCT TCAAGATCAT GACTGACCCA 1680

TTCGTAGGTC GTTTGACATT CTTCCGTGTT TACTCAGGTG TTCTTCAATC AGGTTCATAC 1740

GTATTGAATA CTTCTAAAGG TAAACGTGAA CGTATCGGAC GTATCCTTCA AATGCACGCT 1800

AACAGCCGTC AAGAAATCGA CACTGTTTAC TCAGGTGATA TCGCTGCTGC CGTTGGTTTG 1860

AAAGATACTA CAACTGGTGA CTCATTGACA GATGAAAAAG CTAAAATCAT CCTTGAGTCA 1920

ATCAACGTTC CAGAACCAGT TATCCAATTG ATGGTTGAGC CAAAATCTAA AGCTGACCAA 1980

GACAAGATGG GTATCGCCCT TCAAAAATTG GCTGAAGAAG ATCCAACATT CCGCGTTGAA 2040

ACAAACGTTG AAACTGGTGA AACAGTTATC TCAGGTATGG GTGAACTTCA CCTTGACGTC 2100

CTTGTTGATC GTATGCGTCG TGAGTTCAAA GTTGAAGCGA ACGTAGGTGC TCCTCAAGTA 2160

TCTTACCGTG AAACATTCCG CGCTTCTACT CAAGCACGTG GATTCTTCAA ACGTCAGTCT 2220

GGTGGTAAAG GTCAATTCGG TGATGTATGG ATTGAATTTA CTCCAAACGA AGAAGGTAAA 2280

GGATTCGAAT TCGAAAACGC AATCGTCGGT GGTGTGGTTC CTCGTGAATT TATCCCAGCG 2340

GTTGAAAAAG GTTTGGTAGA ATCTATGGCT AACGGTGTTC TTGCAGGTTA CCCAATGGTT 2400

GACGTTAAAG CTAAGCTTTA TGATGGTTCA TATCACGATG TCGACTCATC TGAAACTGCC 2460

TTCAAGATTG CGGCTTCACT TTCCCTTAAA GAAGCTGCTA AATCAGCACA ACCAGCTATC 2520

CTTGAACCAA TGATGCTTGT AACAATCACT GTTCCAGAAG AAAACCTTGG TGATGTTATG 2580

GGTCACGTAA CTGCTCGTCG TGGACGTGTA GATGGTATGG AAGCACACGG TAACAGCCAA 2640

ATCGTTCGTG CTTACGTTCC ACTTGCTGAA ATGTTCGGTT ACGCAACAGT TCTTCGTTCT 2700

GCATCTCAAG GACGTGGTAC ATTCATGATG GTATTTGACC ACTACGAAGA TGTACCTAAG 2760

TCAGTACAAG AAGAAATTAT TAAGAAAAAT AAAGGTGAAG ACTAATCCGT CCTCACTCTA 2820

GAAGGAAGTC ACTTAGTGGC TTCCTTTTGT CTTTAGAAAA TACCTCTAAA TATGGTAAAA 2880

TAGTAGAAGA ATAATGTGAG GAAAATGAAT GTCAAATAGT TTTGAAATTT TGATGAATCA 2940

ATTGGGGATG CCTGCTGAAA TGAGACAGGC TCCTGCTTTA GCACAGGCCA ATATTGAGCG 3000

AGTTGTGGTT CATAAAATTA GTAAGGTATG GGAGTTTCAT TTCGTATTTT CTAATATTTT 3060

ACCGATTGAA ATCTTTTTAG AATTAAAGAA AGGTTTGAGC GAAGAATTTT CTAAGACAGG 3120

CAATAAAGCT GTTTTTGAAA TTAAGGCTCG GTCTCAAGAA TTTTCAAATC AGCTCTTGCA 3180

GTCCTACTAT AGGGAGGCTT TCTCTGAAGG TCCATGTGCT AGTCAAGGTT TTAAGTCCCT 3240

TTATCAAAAT TTGCAAGTTC GTGCTGAGGG TAATCAGCTA TTTATTGAAG GATCTGAAGC 3300

GATTGATAAG GAACATTTTA AGAAGAATCA TCTTCCTAAT TTAGCCAAAC AACTTGAAAA 3360

GTTTGGTTTT CCAACTTTTA ACTGTCAAGT CGAGAAGAAT GATGTCCTGA CCCAAGAGCA 3420

GGAAGAGGCC TTTCATGCTG AAAATGAGCA GATTGTTCAA GCTGCCAATG AGGAAGCGCT 3480

CCGTGCTATG GAACAACTGG AGCAGATGGC ACCTCCTCCA GCGGAAGAGA AACCAGCCTT 3540

TGATTTTCAA GCGAAAAAAG CTGCAGCTAA ACCCAAGCTG GATAAGGCGG AGATTACTCC 3600

TATGATCGAA GTGACGACAG AGGAAAATCG TCTGGTATTT GAAGGGGTTG TTTTTGATGT 3660

GGAGCAAAAA GTGACTAGAA CAGGTCGTGT TTTAATCAAC TTTAAAATGA CGGACTATAC 3720

TTCAAGTTTT TCTATGCAAA AGTGGGTTAA AAACGAGGAA GAGGCCCAGA AGTTTGACCT 3780

CATCAAGAAG AATTCTTGGC TCCGAGTTCG AGGGAATGTG GAGATGAATA ACTTCACACG 3840

CGATTTGACT ATGAACGTAC AGGATCTGCA GGAAGTTGTT CACTATGAGC GGAAGGATTT 3900

GATGCCAGAA GGTGAGCGTC GGGTTGAGTT TCATGCTCAT ACTAACATGT CGACTATGGA 3960

TGCTTTGCCA GAGGTCGAAG AGATTGTTGC AACAGCTGCT AAGTGGGGAC ACAAGGCGGT 4020

TGCTATCACG GACCATGGGA ATGTCCAGTC CTTTCCACAT GGCTATAAGG CGGCTAAGAA 4080

AGCGGGAATC CAGCTGATCT ATGGGATGGA AGCCAATATC GTGGAGGACC GTGTCCCTAT 4140

CGTCTATAAC GAAGTGGAGA TGGACTTGTC AGAAGCAACC TACGTGGTCT TTGACGTGGA 4200

AACGACGGGA CTTTCAGCTA TCTATAATGA CTTGATTCAG GTTGCGGCTT CTAAGATGTA 4260

CAAGGGGAAT GTTATTGCTG AATTTGATGA ATTTATCAAT CCTGGGCATC CCTTGTCAGC 4320

CTTTACTACA GAGTTAACTG GAATTACAGA TGATCATGTC AAAAATGCCA AACCACTAGA 4380

ACAAGTTTTG CAAGAATTCC AAGAATTTTG CAAGGATACG GTCCTAGTTG CCCACAATGC 4440

TACCTTTGAC GTTGGCTTTA TGAATGCTAA TTATGAGCGG CATGATCTTC CAAAGATTAG 4500

TCAGCCAGTT ATTGATACGC TGGAGTTTGC TAGAAACCTC TATCCTGAGT ATAAACGCCA 4560

TGGTTTGGGG CCTTTGACCA AGCGTTTTGG TGTGGCCTTG GAACATCACC ACATGGCCAA 4620

CTACGATGCG GAAGCGACTG GTCGTCTGCT TTTCATCTTT ATCAAAGAGG TAGCAGAAAA 4680

ACATGGTGTG ACCGATTTAG CTAGACTCAA CATTGATCTA ATCAGTCCAG ATTCTTACAA 4740

AAAAGCTCGG ATCAAGCATG CGACCATCTA TGTCAAGAAT CAGGTAGGTC TAAAAAATAT 4800

CTTTAAGCTG GTTTCCTTGT CTAATACCAA GTATTTTGAA GGAGTGCCAC GGATTCCGAG 4860

AACGGTTCTA GATGCCCATC GAGAGGGCTT GATTTTAGGT TCAGCCTGTT CAGAGGGTGA 4920

640

AGTTTTTGAC GTGGTCGTTT CTCAAGGTGT GGATGCGGCG GTTGAGGTGG CCAAGTATTA 4980

TGATTTTATC GAGGTCATGC CACCGGCTAT CTATGCACCC TTGATTGCCA AAGAGCAGGT 5040

CAAGGATATG GAGGAACTCC AGACCATTAT CAAGAGTTTG ATAGAGGTTG GAGACCGCCT 5100

TGGCAAGCCT GTTCTGGCTA CGGGAAATGT TCACTATATC GAACCGGAAG AAGAGATTTA 5160

TCGTGAAATT ATCGTCCGTA GTTTGGGACA GGGTGCGATG ATTAATCGAA CTATCGGTCA 5220

TGGTGAACAT GCCCAACCAG CACCACTTCC AAAGGCTCAT TTTCGAACGA CTAATGAGAT 5280

GTTGGATGAA TTTGCCTTTT TGGGAGAGGA ACTGGCTCGT AAACTGGTTA TTGAAAACAC 5340

CAATGCCTTG GCAGAAATAT TTGAATCCGT TGAAGTCGTT AAGGGTGACT TGTATACGCC 5400

TTTCATCGAC AAGGCTGAAG AAACAGTTGC TGAGTTGACC TATAAGAAAG CTTTTGAGAT 5460

TTATGGAAAT CCGCTGCCAG ATATTGTTGA TTTGCGGATT GAAAAAGAAT TAACATCCAT 5520

ACTGGGGAAT GGATTTGCTG TGATTTATCT GGCATCGCAG ATGCTGGTGC AACGTTCTAA 5580

TGAACGGGGT TATTTGGTTG GTTCTCGTGG GTCTGTCGGA TCTAGTTTCG TTGCGACCAT 5640

GATTGGGATT ACGGAGGTCA ATCCTCTCTC TCCTCACTAT GTCTGTGGTC AGTGTCAGTA 5700

CAGTGAGTTT ATCACAGATG GTTCGTACGG TTCAGGATTT GATATGCCCC ATAAGGACTG 5760

TCCAAACTGT GGTCACAAAC TCAGTAAAAA CGGACAGGAT ATTCCGTTTG AGACCTTCCT 5820

TGGTTTTGAT GGGGATAAGG TTCCTGATAT TGACTTGAAC TTCTCGGGAG AAGATCAGCC 5880

TAGCGCCCAC TTGGATGTGC GTGATATCTT TGGTGAAGAA TATGCCTTCC GTGCGGGAAC 5940

GGTTGGTACG GTAGCTGCCA AGACTGCCTA TGGATTTGTC AAAGGTTACG AGCGAGATTA 6000

TGGCAAGTTT TATCGTGATG CAGAAGTAGA ACGCCTCGCT CAAGGAGCGG CGGGTGTCAA 6060

GCGGACAACA GGCCAACACC CGGGGGGAAT CGTTGTTATT CCGAACTACA TGGATGTCTA 6120

CGATTTTACG CCTGTCCAGT ATCCAGCAGA TGATGTCACG GCTGAATGGC AGACCACTCA 6180

CTTTAACTTC CACGATATCG ATGAGAACGT CCTCAAACTC GATGTACTGG GACATGATGA 6240

TCCGACTATG ATTCGAAAAC TTCAGGATTT GTCTGGTATT GACCCTAATA AAATTCCTAT 6300

GGATGACGAA GGCGTGATGG CACTCTTTTC TGGGACTGAT GTGCTAGGGG TAACACCTGA 6360

ACAAATTGGA ACGCCTACGG GTATGTTGGG GATTCCAGAG TTTGGAACAA ATTTCGTACG 6420

TGGAATGGTA GACGAAACCC ATCCGACAAC CTTTGCGGAA TTGCTTCAGC TGTCTGGTCT 6480

GTCCCACGGT ACTGATGTTT GGTTGGGGAA TGCTCAGGAT CTGATTAAGC AAGGAATAGC 6540

GGACCTATCG ACTGTTATCG GTTGTCGGGA CGACATCATG GTTTACCTCA TGCATGCGGG 6600

TCTGGAACCT AAGATGGCCT TTACCATTAT GGAACGGGTA CGTAAGGGTT TGTGGCTAAA 6660

GATTTCAGAA GAGGAGAGAA ATGGCTATAT CGAAGCAATG AAGGCTAATA AGGTGCCAGA 6720

GTGGTATATC GAATCCTGTG GGAAAATTAA GTACATGTTC CCTAAGGCCC ATGCGGCAGC 6780

CTACGTTATG ATGGCCTTGC GTGTAGCTTA CTTCAAGGTT CACCATCCTA TTTATTACTA 6840

CTGTGCTTAC TTCTCCATTC GTGCTAAGGC TTTTGATATC AAGACCATGG GTGCGGGCTT 6900

GGAGGTCATC AAGCGCAGAA TGGAAGAAAT CTCTGAAAAA CGGAAGAACA ATGAAGCCTC 6960

TAATGTGGAA ATCGATCTCT ATACAACTCT TGAGATTGTC AATGAGATGT GGGAACGAGG 7020

TTTCAAGTTT GGTAAATTAG ATCTCTACTG TAGTCAGGCG ACAGAGTTCC TCATCGACGG 7080

GGATACCCTT ATCCCACCAT TTGTAGCAAT GGATGGTCTG GGAGAGAACG TTGCCAAGCA 7140

ACTGGTGCGG GCGCGTGAAG AGGGAGAATT CCTCTCTAAA ACAGAACTAC GCAAGCGTGG 7200

TGGACTCTCA TCAACCTTGG TTGAAAAGAT GGATGAGATG GGTATTCTTG GAAATATGCC 7260

AGAGGATAAC CAGTTGAGTT TGTTTGATGA GTTGTTTTAA AAAATTGCTT AATAATCTAT 7320

TAAAAGAGGC TAACGTATAT CCAATAGATT TACATTAGCT TTCTTTTTTG TTAAAATAGT 7380

CTATGGAAAG AGGGTGAGAG TATGTCAAAG ATGAGTATAA GCATCCGTCT GGATAGTGAG 7440

GTTAAGGAGC AGGCCCAACA GGTGTTTAGT AATCTGGGAA TGGATATGAC AACAGCTATT 7500

AATATTTTCC TTCGTCAGGC AATTCAATAT CAGGGATTAC CTTTTGATGT TAGACTAGAC 7560

GAAAATCGGA AGTTGCTCCA AGCGTTAACG GATTTAGACC AAAATCGTAA TATGAGCCAG 7620

TCTTTTGAAT CAGTCTCAGA TTTGATGGAG GACTTACGTG CTTAAGATTC GTTATCATAA 7680

ACAGTTTAAA AAAGATTTTA AGTTGGCTAT GAAGCGTGGT TTGAAGGCAG AATTATTAGA 7740

AGAAGTTTTG AATTTTCTGG TTCAAGAAAA AGAACATCCT GCCAGAAATC GTGATCATTC 7800

ATTGACGGCA TCCAAGCATT TTCAAGGAGT TCGTGAATGC CATACCCAGC CAGATTGGCT 7860

TTTGGTTTAT AAAGTAGACA AGTCGGAATT GATTTTAAAT TTGCTGAGGA CAGGCAGTCA 7920

CAGTGATTTA TTTTAATCTA TTTTAAGGGG GTTCTCATGA AACTAAGAAT ATTTGCGGAA 7980

GATAAGCCGG CTAAGAAGGT ATTTGAATAT CAATTAGAAC TTGCTGATCG TACAATTCTT 8040

CTATCGACAG CACTCTTGTC AGGTGCTATT GCTTTAGCAG GAATCTTTTC TGCTTTGAAA 8100

GAAAAATAAA AATAGAAAAG AGAAAACAGA ATGGTTTTAC CAAATTTTAA AGAAAATCTA 8160

GAAAAATATG CGAAATTGTT GGTTGCGAAC GGAATTAACG TGCAACCTGG TCACACTTTG 8220

GCTCTCTCTA TTGATGTGGA GCAACGTGAA TTGGCACATC TAATCGTGAA AGAAGCTTAT 8280

GCCTTGGGTG CGCATGAGGT CATCGTTCAG TGGACAGATG ATGTGATTAA CCGTGAGAAA 8340

TTCCTCCATG CCCCGATGGA GCGTTTGGAC AATGTGCCAG AATACAAGAT TGCTGAGATG 8400

AACTATCTCT TGGAGAATAA GGCTAGCCGT CTTGGAGTTC GTTCATCTGA TCCAGGTGCC 8460

TTGAACGGAG TGGACGCTGA CAAGCTTTCA GCTTCTGCTA AAGCTATGGG ACTTGCCATG 8520

AAGCCTATGC GTATCGCAAC TCAATCTAAC AAGGTTAGCT GGACTGTAGC AGCTGCAGCA 8580

GGACTTGAGT GGGCTAAGAA AGTCTTCCCA AATGCTGCGA GCGACGAAGA AGCAGTTGAT 8640

TTCCTTTGGG ACCAAATTTT CAAAACTTGC CGTGTCTACG AAGCAGATCC TGTTAAGGCT 8700

TGGGAGGAAC ATGCAGCCAT TCTCAAGAGC AAGGCCGATA TGCTTAATAA GGAGCAATTT 8760

TCAGCCCTTC ACTACACAGC GCCAGGAACA GATTTAACAC TTGGTTTGCC AAAGAACCAC 8820

GTTTGGGAAT CAGCTGGTGC TGTCAATGCA CAGGGCGAAG AATTCTTGCC AAATATGCCA 8880

ACAGAAGAGG TCTTCACAGC GCCTGACTTC CGTCGTGCAG ATGGTTATGT CACTTCTACA 8940

AAACCGCTTA GCTACAACGG AAATATCATT GAAGGCATTA AGGTGACCTT TAAGGATGGA 9000

CAAATCGTAG ATATCACTGC TGAGAAGGGT GATCAGGTTA TGAAAGACCT TGTCTTTGAA 9060

AATGCGGGTG CGCGTGCCTT GGGTGAATGT GCCTTGGTAC CAGATCCAAG TCCAATTTCT 9120

CAGTCAGGCA TTACCTTCTT TAACACCCTT TTCGATGAAA ATGCGTCAAA CCACTTGGCT 9180

ATCGGTGCAG CCTATGCGAC TAGCGTTGTT GATGGAGCGG AGATGAGCGA AGAGGAGCTT 9240

GAAGCTGCAG GGCTTAACCG TTCAGATGTT CACGTAGACT TTATGATTGG TTCTAACCAA 9300

ATGGATATCG ATGGTATTCG TGAGGATGGA ACGCGGGTAC CTCTTTTCCG TAATGGGAAT 9360

TGGGCAAATT AAGGAGATAA TATGTTAGGA AGTATGTTCG TTGGTCTCCT AGTGGGATTT 9420

TTAGCAGGTG CTATGACCAA TCGTGGAGAG CGAATGGGAT GTTTTGGAAA AATGTTTCTC 9480

GGTTGGATCG GAGCCTTTCT AGGTCACTTG CTCTTTGGAA CTTGGGGGCC AGTTTTATCA 9540

GGAACAGCTA TTATCCCAGC GATTTTAGGA GCCATGATTG TTTTAGCTAT TTTTTGGAGA 9600

CGAGGAA 9607

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14231 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

CTACAAGATA ATTCCAGCTA TAACATCCGC TATAATAGTA AGAGCGAGCT CTATGATAAG 60

GCTCATTAGT TTCACCTCCT CTCACGAACC CATAGGAACG TAATCGGTAA CCGATGACAA 120

AAATAGTATA CCACAATACA TTTAGATCAT CAAGGTCACT TAATTCTTGA AATATCAGAT 180

CTAAGAGAAA AATCTTTAAA ATCAGAAAAA CGCATAATAT CAGGTGTGCA AAAACTTGAT 240

ACTATGCGTT TTATTGTGGG AAGGTTTACT CCATTTTCTC CTGAAATTGA GTTTTTGTCC 300

AGCCTCTGTT TTTAGGGTTG CTAAGAAAAT AATGTCATGT GGTGAATATT TGTAAATCAG 360

TCAGCAGACA GAACGATACT CTTCGAAAAT CTCTTCACAT CATGTCAGCT TCGTCTTTCC 420

GTATATATGT GACTGACTTC ATCAGTTCTA TCTACAACCT CAAAACAGTG TTTCGAGCTG 480

ACTTGATCAA TTTTCAAATC TGTACTTTGA GCAAGCTGAG ACTAGCTTCC TATTTGATTT 540

TCATTGAATA TCAGAAACCC ATTCTCCATC AAATAATTCG ACTGCGTCTA ATAATTTTTG 600

ATCTGGCACG GTGTCTGAAA TAAAGGTTGT GTATTTGGAG AGGGGATTAA TTTTAAAAAA 660

TCCAGTCTTG TAAAATTTAG AACTATCAAT CAGTAAGATG GTTTCATGGG CTTTGTCAAT 720

AATATTCTTT TTTGAAATAG CTTGGCTGAG AGAAGCTTCA TAAACATATT GGTCATCAAT 780

ACCTCTTGCT GAACAAAATG CTAAATCGAT ATTAAAATGA TCTAATAAAG AATTTTCCTT 840

ATCATAGTTG ACCACGGAAC AGGATTGATG TTTGACCTCG CCAGATGTGA TAAAGATTTT 900

GGAGCTATCT TTAACAGTTT CAGATAGGGT TTGTGCAGTA TGTAAACCAT TTGTAAAAAT 960

AATCAAATTA TCAAGTTCAG AAAGATAGGG ACAGAGTTCG TAGACAGTAG TACTAGAATC 1020

TAGATAGATA CACATACCAG ACCGAATAAA GTCTTTAGCG AGACTAGCGA TTAGTCTTTT 1080

TTGCCTAGTA CTTTCTCCTT CACGTATTTG ATGAGAAAGT TCAATTGTGT TCATAGAGGA 1140

CAGGGTCACG TATCCGTGCT TTCTTTTGAT AAGACCTTGA TTTTCTAAGA AAATTAAATC 1200

ACGACGTAAG GTACTTGTGC TGGAGAAAGT GATTTCTGCC AGCTCTTTTA CGGCAATTCT 1260

TTTTTTCTTT TTGATAATTT CAATCAATTC AAGTACACGT TCATCTTTTA TCATAAGCTC 1320

CTCCTAATTT ATCATTTCAA CTATATTATA GCACAAATTG GAGGAATTTG AATTATTTTT 1380

ATGAATATTG GGTTAACATT TGAACATTAT TCAAGTAAGC GTTCACATAT TGAAAAAATA 1440

AAACGTGGGG ATTATAATAA AGTTAATCmA GGACGAAGAG AGAAGAAAAA TGGAAGCGGT 1500

TTTAGCAATA GATTTAGGTG CGACTTCTGG AAGAGCAATC GTTGGTTACC TTTCTGAAAA 1560

TAAACTAGTA ATGGAAGAAA TAAATCGCTT TTCTAATCTA CCTATTAGAG TAAAAGGGCA 1620

TTTATCTTGG GATATTGACT TTCTACTAGC TAAAATTCTT GAAAGTATCC GCTTGGCTAA 1680

TACTAGTTAC AAGATTTTAT CTATCGGTAT TGACACATGG GGAGTTGATT TTGGACTGAT 1740

TGATAATGAA GGTAAGCTGT TATTACAACC TGTTCATTAT CGTGATGAAA GAACAAAGGG 1800

AGTGTTAAAG GAAATATCTG AAATGACTGA ATTAGAAAAA CTGTATTCAG AGACAGGAAA 1860

TCAGATTATG GAGATAAATA CCTTGTTTCA ACTCTTTAAG GCACGTCAAG AATCTCCTGA 1920

CTCTTTCTAT AAGACCAATA AGATTCTTTT AATGCCAGAT TTGTTTAATT ATCTCTTGAC 1980

AGGTAAGTTT GCTACAGAAA AAAGCATTGC TTCAACAACT CAATTATTTG ATCCTAGGAG 2040

TCAAAATTGG AATCAGAATA TCTTAAAACT ATTTGAATTG GATTCATCTT TACTTCCTGA 2100

AATTGTTTCA GAGGGAAATG TTCTTGGAAG GATAAAAGAG GAGTATGGTT TAGGCGATAT 2160

TCCTGTTGTG AATGTTTGTA GTCATGATAC AGCAAGCGCG ATTGTCTCAG TACCTAAGAC 2220

AGAAGGTAGT TTATTTATTT CATCAGGTAC TTGGTCTTTG GTTGGAGTGG AACTTACTTC 2280

ACCGATTCTT ACTACCGAAT CCTTCAGTTA TGGATTTACA AATGAAGTCG GTAAAGATGG 2340

AGTGATTACA TTTCTGAAGA ATTGTACAGG GTTGTGGATC ATAGAGGAAC TAAGACGTTC 2400

ATTTGAACGA AGAGGGAAAG CCTATTCTTT TGATGATATT AGGACAATGG TGGAGAAAGA 2460

AAAAGAAAAT CTTCCTCTGA TTGATACTGA ATCAACTGAA TTTGCAACAG AATCTGATAT 2520

GCACAAGACT TTGACAGAAT ATCTAGCTTA TCATCATGAA ACTAGAGAGT GGACAGATGG 2580

ACAACTATTT AAGATTGTTT ATGAAAGCCT AGCTGAAACG TATAGGAAAG CGATAGAGTT 2640

ACTAGAAGAA CTAACTCATA AGGTTTATAA GAGGATATAT GTGATTGGAG GAGGTGCTAG 2700

AGCCAGTTAC TTTAACCAAA TGATTGCTGA TAGAACTGGT AAAGAGGTTC TTACAGGTTT 2760

GACTGAGGCT ACAGCTGTGG GGAATATTGT TGTGCAGCTC ATAGCTATGG GACAATTAAA 2820

AGGGATGGAA GAGGCTCACC ATGTTATTGA GGAGTTTCTA CAATTAGAGA GTTATTACTC 2880

CCAAAAGAAT TAAAAAGATT GAGAGTTTGT AAATTTGCCT CCCTCCCCCT TCTTAGCTTT 2940

TGTGCAGGAA GGGGGGATAA TTGGTGAATT GAAAAATATT TAGTGTTTTG ATATGAGGAG 3000

GACAAGGATG TCAGATGTAA AACAAGAATT AATTAAATAT GGTAAGAAGC TAGTAGAAAC 3060

AGATTTGACG AAAGGAACAG GTGGGAATCT CAGCGTTTTC GATCGTGAAA AACAATTGAT 3120

GGCAATTACC CCGTCGGGTA TTGATTTCTT TGAAATCAAA GAATCCGATA TTGTAGTGAT 3180

GGATATTAAT GGAAATGTTG TAGAGGGAGA ACGCTTGCCA TCTAGCGAAT GGTATATGCA 3240

TTTGATTCAA TATCAAACTC GTGATGATAT CGATGCAATT ATCCATGCTC ATACAACTTA 3300

TGCAACAGTA TTAGCTTGTC TCAGAGAACC ACTTCCAGCG AGTCATTATA TGATTGCAGT 3360

GGCAGGGAAA GATGTTCGGG TAGCTGAGTA TGCAACATAT GGCACGAAAG AATTGGCTGT 3420

GAATGCAGCT AAAGCAATGG AAGGTCGTAG AGCAGTTTTA CTAGCGAATC ATGGAATTTT 3480

AGCAGGTGCA CAAAATTTAT TGAATGCATT TAATATTGTT GAAGAAGTTG AATATTGTGC 3540

AAAAATTTAT TGTTTAGCTA AGAATTTTGG AGAGCCAGTA GTTCTTCCTG ATGAGGAGAT 3600

GGAATTGATG GCAGAAAAAT TTAAAACATA CGGTCAGAGA AAATAGGGAG GATATTAATG 3660

TTAAAACATA TACCGAAAAA TATTTCTCCA GATTTATTGA AGACTTTAAT GGAAATGGGA 3720

CATGGAGATG AAATAGTATT AGCTGACGCG AATTATCCTT CTGCCTCATG TGCAAATAAG 3780

CTAATTCGTT GTGATGGTGT AAATATTCCA GAATTATTAG ATTCCATTCT GTATTTAATG 3840

CCATTAGATA GTTACGTCGA TAGTTCAATT CAGTTTATGA ACGTTGTTTC GGGTGATGAT 3900

ATTCCTAAGA TATGGGGTAC CTATAGACAG ATGATTGAAG GTCATGGTAC AGATCTTAAA 3960

ACGATTACTT ATCTTAGAAG AGAAGACTTT TATGAACGTA GTAAGAAAGC TTATGCTATT 4020

GTTGCTACAG GAGAAACTTC ACTTTATGCT AATATTATCC TTAAGAAAGG AGTAGTTGTT 4080

GAAAGAGAAA ATGTTCAATA GAGGAATTTT AGTTGCCAGT CATGGTAATT TTGCTAGCGG 4140

AGCTCTCATG ACCGCAGAAA TGTTTGTTGG TGAGACAACA AATGATAGAG TTAGGACATT 4200

AGGTTTGATG CCTGGAGAGA ATATTGTAGA GTTTGAGCAT TATTTTAAAA ATCAAGTGGA 4260

TGAACTGTTA GACTCAAATC AAGAGGTTAT CGTTTTGACT GACTTGATTG GAGGAAGTCC 4320

TAATAATGTG GCTTTGTCAC GGTTTTTAAA TTTGGATTCA GTTGATATTG TAACAGGGTT 4380

TAATATCCCT CTCCTAGTGG AATTAATATC AAGTTATGAT TCAAAAATCA ATTTAGAAGA 4440

AATTGTTCAC AATGCTCAAA ATAGTTTGTT TAATGTTAAA CAACAACTTA ACGTAGAGGA 4500

GGAAGAAGAT TTATGTCTAT AGAGTTTGTT CGTATTGATG ACCGTCTGGT ACATGGTCAA 4560

GTTGTCACTA CGTGGCTAAA AAAGTATGAT ATTGAGCAAG TTATCATTGT TAATGATCGC 4620

ATCTCAGAAG ATAAAACACG ACAATCTATT TTAAAGATTT CTGCACCGGT AGGTTTAAAA 4680

ATTGTTTTCT TTAGTGTAAA ACGGTTTGTG GAAGTTTTAA ACTCTGTGCC AATAAAAAAG 4740

AGAACAATGC TGATATATAC AAATCCAAAA GATGTGTATG ATTCTATTGA AGGAAATTTA 4800

AAATTGGAGT ACCTCAATGT AGGACAGATG AGTAAAACGG AGGAAAATGA AAAGGTAACG 4860

GGAGGTGTAG CTCTAGGTGA AGAAGACAAA TATTATTTTA AGAAAATAGT TGATAAGGGA 4920

ACGAGAGTTG AAATTCAAAT GGTTCCTAAT GATAAAGTTA CAATGTTGGA AAAATTTTTA 4980

TAAAAATAAT TTAAGGAGGT ACAGTATATG CTATTCACAC AAGCATTACT GGTGACATTA 5040

GTTGGGATTA TTGCCACTAT TGACTATAAT GGACCGTTAT TTATGATTCA CCGTCCGTTA 5100

GTTACAAGTG CAATGGTTGG CTTAGTATTA GGAGATTTCA CCCAAGGTGT TCTTATTGGT 5160

TCAGCTCTTG AATTAACTTG GCTCGGTGTA ACAGGTATTG GAGGTTATAC TCCACCAGAT 5220

ACTATTTCAG GTGCGATTAT TGGTACTGCA TTTGGTATTT TATCTGGTCA AGGAGAAACT 5280

GCTGGTATCG CTATAGCAGT TCCAATTGCA GTTGCTACCC AACAGTTGGA TGTTCTTGCA 5340

AAAACTTTAG ATGTTTATTT TGTGAAAAAA GCTGATAATG ATGCTAAAAA CGGAGATTAT 5400

TCAAAGATCG GTTTTTATCA TTATTCAAGT TTGGTTTTAA TCACGTTATT TAAAATTGTA 5460

CCAATTTTCC TAGCTATTAT GCTTGGAGGG GAATATGTGG CAGACTTGTT TGCTAAGGTT 5520

CCACCAATCG TTATGCAGGG ACTTAACTCT GCAGGTGCTT TACTACCTTC AATTGGTTTT 5580

GGTATGCTTT TAAATATGAT GCTCAAGAAA AATATGTGGG TATTCTTGTT GATTGGATTC 5640

ATTTGTTCTG TGTATGGAGG AATGTCAACC ATTGGGATCT CACTAGTTGG TATTGCGGTA 5700

GCATACTTCT ACGATATGAT TGGAAGCAAA CCACAAGAAA CAACTTCAAG TAGTGATGTT 5760

GAGGAGGATC TTGATCTATG ATGAATAATA AAGTAACTAA AGTTGAACTT AAAAAAGTTT 5820

TCAAACGAAG TTTTATGTAT GGTTCTTCAT GGAACTATGA GAGAATGCAG AACCTAGGTT 5880

TTCTATATAC AATTCTTCCA GTATTGAAAA AACTATACCC AGACAAAGAT TCAGCTTCTC 5940

CTGCAATGAA ACGTCACCTT GAGTTTTTCA ATACTCATCA AACAGCGGCA CCATTTATTC 6000

TTGGAGTTAC TTCCGCTATG GAAGAACAAG AAGGAAATGA AGGTGCAGCT TCAATTACTG 6060

GTATTAAAGT TGGCTTGATG GGGCCACTGG CTGGTCTAGG AGATAGTTTG TTCTGGCTGA 6120

CACTAGTTCC TATCTGTTTT AGTATTGGTG CGTCTTATTC TAAAGACGGC GGTGCTTTAG 6180

GTATCTTTAT CGCCTTAATA TTGTTTAATA TTATTAATAT TCCTGTTAAA TATTTCGGTT 6240

TGAAATATGG GTATACTAAG GGTTCTAGTC TTATCCAAGA AAATAATACA AAAGGAACAT 6300

TGAATCGCGT TACGAGTATG GCGACAGCAT TAGGGCTAGT ACTAGTGGGT GGTTTGATTC 6360

CATCAATGGT TGGTATTAAT TTTGGATTAG AATTTAAGCA GGGGGAACTT GTTATTTCTG 6420

TTCAAGAAAT GATTACAAAA TTAATTCCAG GATTTATCCC TATGGCTTTG ACTTTATTAA 6480

TGTGTAAATT AATTAGAAAA GGAAAGAATC CGGTTGTACT AATCTTTAGT GTTATGGCTA 6540

TTGGAGTTAT TCTAGTTGTT TTAGGAATTT TGAAGTAGTA GAAAGTGTGG AGGTGGTATT 6600

TGGGATATCA CCTCCATTTT GGAAGAGAGG TAAAGAGTGA AATTATGGTA TAAGAAAGCT 6660

GCCGCAAATT GGAATGAAGC CTTGCCGATT GGGAACGGTC ATTTAGGTGG TATGATTTAT 6720

GGTTCAGCTA CAAAAGAATG TATTCAACTA AACGATGAGA CTATTTGGTA TAGAGGAAAG 6780

TCAGATAGAA ATAATCCAGA CTCACTATTG CATCTTAAAA AAATTCGGGA ATATCTTTTA 6840

GATGGAGAAA TTCAGAAAGC CGAAGAATTG ATAAAGTTAA CAGTGTTTGC TACCCCAAGA 6900

GATCAAAGCC ACTATGAATT ACTTGGGGAA CTTTACATTG AGCATATAGA TATTCAGTCT 6960

TGTGCTCTTT CATTGTATGA AAGAGAGCTA GATTTAGATA CAGCTATTTC TAATGTTGTG 7020

TTTGAGCCTA ATAGTTGTAA TTTACAAATA AAAAGAGAAT ATTTTACGAG TTTTAATAAG 7080

AATATTTTAT GTTGCCGTAT AGTGTCATCA GTTCAAAACA CATTAAATTT AAACATTAAT 7140

TTGGGTAGAA ATAAACGGTT TAATGACGAA GTATCTAAAC TGGATTCAAG TACAATTTTA 7200

ATGTCGGCCT CTGCTGGAGG TAGAAAAGGT GTTCAGTTTA AAGTAGTATG TCATTCTAAG 7260

GTTACGGATG GTGAAGTAAG TGTATTGGGA GAGACAATAG TTATTCGGAA TGCTACAGAG 7320

GTATTTCTTT ATCTCAAATC AATGACGGAT TATTGGGGAA ATATAGATAT TTCTTCTCTT 7380

CAGGGAGAAT TTAGTAGTAT TGATTACTTT ACAGAAAAAG ATGAACATGT AAAAAAATAT 7440

CAGGAGCAAT TTAATAGAGT TGATTTTAAA CTAGACTATA GTAAAGGTTG TCTTAGCATT 7500

CCAACGAATC TACTTCTTGA AAACACTAAA AAGTATAGTA ACTACTTGAC TAACTTGTTA 7560

TTTCATTATG GAAGATATCT GTTAATATCG TCTAGTCAAC CGAATGGTTT ACCTGCCAAT 7620

CTTCAAGGAA TATGGTGTGA TGAATTAAAT CCAATTTGGG GTTCTAAATA TACGATTAAT 7680

ATTAATACTC AAATGAATTA TTGGATGGTA GGTCCATGTG ATTTACCAGA AGTAGAATAT 7740

CCATTATTTG ATATGCTCGA AAGAATGAGA GAACCGGGAA GACTAACCGC TAAGAAAATG 7800

TATGGAGCTA GAGGTTTTAC AGCACATCAT AATACGGATG GTTTTGGCGA TACGGCTCCC 7860

CAATCTCATG CCATGGGGGC TGCAATTTGG GTATTAACTA TTCCATGGTT ATGTACTCAT 7920

ATTTGGGAAC ACTATTTATA TTTCCAAGAT GAGCGTATTC TTACGGAACA TTTTGAAATG 7980

ATAAAAGAAG CATTTCTTTT CTTTGAAGAT TATTTATTTG AGGTGGATGG CTACTTGATG 8040

ACAGGTCCAA GTGTCTCACC GGAAAATAAA TATCGCTTAA AAAATGGTAT TGAAGGAAAT 8100

GCTTGTCTAT CATCTACAAT TGATAATCAA ATTCTAAGAT ATTTTTGTGA TTCATGCATT 8160

GGCATTGCAA AACAATTAGG AGACAATTCG GATTTTATTA GTCGTGTGAA GGAGTTAAAA 8220

AAGAAACTAC CTAAAACAAA AATAGGTAGT AATGGGCAAA TCCAAGAATG GTTAGAAGAT 8280

TATGAAGAAG TAGAGCCTGG GCATAGACAC ATTTCACCTC TATTTGGGCT TTATCCTTAT 8340

AATGAGATTG ATATTCATAA AACTCCGGAA TTAGCAGAAG CAGCTAAAAT CACTATCAAT 8400

AGGAGATTAT CAAACGCTAA TTTTTTATCT TCACAGGAGA GGGAGCAAGC GATTAATAAT 8460

TGGTTAGTAA GTGGTTTGCA TGCTAGTACA CAAACAGGTT GGAGTGCTGC ATGGCTGATT 8520

CATTTTTTTG CGAGACTATA TCAAGGTGAA CCTGCTTATA ACCAGATTAA TGGTTTGTTA 8580

AATAATGCGA CTCTTGGCAA TTTATTTCTT GACCATCCAC CATTTCAAAT TGATGGTAAT 8640

TTAGGTTTGG TGAGTGGAAT TTGTGAATTA TTAGTACAGA GCCATCATAA TTGGTTATCA 8700

CTAATTCCAG CTTTACCTTC TGCTTGGTCA GAAGGAGAAG TGAAAGGTTT CAGAGTAAGA 8760

GGAGGATATA AGGTATCGTT TGCTTGGAAA AATGGGGATA TAACATTCCT AAAATTGGAA 8820

GGAGGAAACA AAGATCAAAA AGTAAGAGTA AGAATATATG GCAAAAATAC TGATGTACAA 8880

AATATTGAAT TGGTATTTAA TTCAGAAAAA ATTATTGAGT TAAATTTTTA GGTATAAGTC 8940

ATGAATAAAG AAAAAATAAA AAGAAAATTA ATCACAATAT TGTTTGTATG TATTGGGATG 9000

TTATGTTTTG GATTGTTAGC AGGAGTTAAG GCTGATAATC GTGTTCAAAT GAGAACGACG 9060

ATTAATAATG AATCGCCATT GTTGCTTTCT CCGTTGTATG GCAATGATAA TGGTAACGGA 9120

TTATGGTGGG GGAACACATT GAAGGGAGCA TGGGAAGCTA TTCCTGAAGA TGTAAAGCCA 9180

TATGCAGCGA TTGAACTTCA TCCTGCAAAA GTCTGTAAAC CAACAAGTTG TATTCCACGA 9240

GATACGAAAG AATTGAGAGA ATGGTATGTC AAGATGTTGG AGGAAGCTCA AAGTCTAAAC 9300

ATTCCAGTTT TCTTGGTTAT TATGTCGGCT GGAGAGCGTA ATACAGTTCC TCCAGAGTGG 9360

TTAGATGAAC AATTCCAAAA GTATAGTGTG TTAAAAGGTG TTTTAAATAT TGAGAATTAT 9420

TGGATTTACA ATAACCAGTT AGCTCCGCAT AGTGCTAAAT ATTTGGAAGT TTGTGCCAAA 9480

TATGGAGCGC ATTTTATCTG GCATGATCAT GAAAAATGGT TCTGGGAAAC TATTATGAAT 9540

GATCCGACAT TCTTTGAAGC GAGTCAAAAA TATCATAAAA ATTTGGTGTT GGCAACTAAA 9600

AATACGCCAA TAAGAGATGA TGCGGGTACA GATTCTATCG TTAGTGGATT TTGGTTGAGT 9660

GGCTTATGTG ATAACTGGGG CTCATCAACA GATACATGGA AATGGTGGGA AAAACATTAT 9720

ACAAACACAT TTGAAACTGG AAGAGCTAGG GATATGAGAT CCTATGCATC GGAACCAGAA 9780

TCAATGATTG CTATGGAAAT GATGAATGTA TATACTGGGG GAGGCACAGT TTATAATTTC 9840

GAATGTGCCG CGTATACATT TATGACAAAT GATGTACCAA CTCCAGCATT TACTAAAGGT 9900

ATTATTCCTT TCTTTAGACA TGCTATACAA AATCCAGCTC CAAGTAAGGA AGAAGTTGTA 9960

AATAGAACAA AAGCTGTATT TTGGAATGGA GAAGGTAGGA TTAGTTCATT AAACGGATTT 10020

TATCAAGGAC TTTATTCGAA TGATGAAACA ATGCCTTTAT ATAATAATGG GAGATATCAT 10080

ATTCTTCCTG TAATACATGA GAAAATTGAT AAGGAAAAGA TTTCATCTAT ATTCCCTAAT 10140

GCAAAAATTT TGACTAAAAA TAGTGAGGAA TTGTCTAGTA AAGTCAACTA TTTAAACTCG 10200

CTTTATCCAA AACTTTATGA AGGAGATGGG TATGCTCAGC GTGTAGGTAA TTCCTGGTAT 10260

ATTTATAATA GTAATGCTAA TATCAATAAA AATCAGCAAG TAATGTTGCC TATGTATACT 10320

AATAATACAA AGTCGTTATC GTTAGATTTG ACGCCACATA CTTACGCTGT TGTTAAAGAA 10380

AATCCAAATA ATTTACATAT TTTATTGAAT AATTACAGGA CAGATAAGAC AGCTATGTGG 10440

GCATTATCAG GAAATTTTGA TGCATCAAAA AGTTGGAAGA AAGAAGAATT AGAGTTAGCG 10500

AACTGGATAA GCAAAAATTA TTCCATCAAT CCTGTAGATA ATGACTTTAG GACAACAACA 10560

CTTACATTAA AAGGGCATAC TGGTCATAAA CCTCAGATAA ATATAAGTGG CGATAAAAAT 10620

CATTATACTT ATACAGAAAA TTGGGATGAG AATACCCATG TTTATACCAT TACGGTTAAT 10680

CATAATGGAA TGGTAGAGAT GTCTATAAAT ACTGAGGGGA CAGGTCCAGT CTCTTTCCCA 10740

ACACCAGATA AATTTAATGA TGGTAATTTG AATATAGCAT ATGCAAAACC AACAACACAA 10800

AGTTCTGTAG ATTACAATGG AGACCCTAAT AGAGCTGTGG ATGGTAACAG AAATGGTAAT 10860

TTTAACTCTG GTTCGGTAAC ACACACTAGG GCAGATAATC CCTCTTGGTG GGAAGTCGAT 10920

TTGAAAAAAA TGGATAAAGT TGGGCTTGTT AAAATTTATA ATCGCACAGA TGCTGAGACT 10980

CAACGTCTAT CTAATTTTGA TGTGATTCTA TATGACAATA ATAGAAACGA AGTTGCTAAG 11040

AAACATGTTA ATAATTTGTC GGGTGAATCT GTTAGTCTAG ATTTCAAAGA AAAAGGAGCA 11100

AGGTATATTA AAGTTAAATT ACTAACGAGT GGAGTGCCTT TGAGTTTAGC AGAAGTAGAG 11160

GTTTTTAGAG AATCAGATGG TAAGCAATCT GAAGAGGATA TAGATAAAAT AACAGAAGAT 11220

AAAGTAGTCT CTACAAATAA GGTAGCTACT CAAAGTTCAA CCAATTATGA GGGTGTAGCT 11280

GCTTTAGCAG TTGATGGTAA TAAAGATGGA GATTACGGAC ATCATTCGGT GACTCATACT 11340

AAGGCAGATT CTAACGCTTG GTGGCAGGTC GATCTGGGAG AAGAGTTTAC GGTTTCTAAA 11400

GTTGATATTT ATAATAGAAC AGATGCCGAA CCTCAGCGTT TATCTAATTT TGATGTTATT 11460

TTTCTATCTT CATCAGGAGA AGAAGTTTTT AGAAGACATT TTGATAAAGT AGTTGATGGT 11520

TTGTTATCTT TAAAAGTACC TTCTGTAGGG GCTAAGCTAG TCAAAATAGA ATTAAAATCA 11580

GCAGCTATTC CGTTAAGTTT AGCGGAAGTT GAAGTCTATG GTTCAAAGAG AACTCCGAAG 11640

AAACTTTCTA ATATTGCATT AACAAAAGAA ACTCGACAGA GTTCAACGGA TTACAATGGT 11700

TTTTCTCGTC TAGCAGTTGA TGGAAATAAA AACGGAGATT ATGGTCATCA TTCAGTGACT 11760

CATACCAAAG AAGATTCTCC TTCATGGTGG GAGATAGATT TAGCACAAAC CGAAGAATTA 11820

GAAAAGTTAA TTATTTATAA TAGAACAGAT GCTGAAATTC AGAGATTATC AAATTTTGAT 11880

ATTATTATAT ATGATTCAAA TGATTATGAA GTTTTTACAC AACATATTGA CAGTTTAGAA 11940

AGCAATAATC TATCCATAGA CTTAAAAGGA CTGAAGGGAA AAAAGGTTAG AATTTCTTTG 12000

AGAAGCGCAG GAATTCCTTT AAGTTTAGCA GAGGTAGAGG TTTATACTTA TAAGTAATTT 12060

TAAAAATTAT CACCCAGGCT ACCGTAAATA TAATGGAGAT GGTAGTATGA AAGAAACAGA 12120

AAAATAAGAG GAAAATAGTA TGATTCAACA TCCACGTATT GGGATTCGTC CGACTATTGA 12180

TGGTCGTCGT CAAGGTGTAC GCGAATCACT TGAAGTGCAA ACAATGAACA TGGCTAAAAG 12240

TGTGGCAGAT TTGATTTCAA GCACATTGAA ATATCCAGAT GGGGAACCTG TGGAATGCGT 12300

GATTTCTCCA TCTACTATTG GCCGTGTACC AGAGGCTGCA GCTTCCCATG AGTTGTTTAA 12360

AAAATCAAAT GTTTGCGCAA CAATTACAGT TACACCATGC TGGTGTTATG GTAGTGAAAC 12420

TATGGATATG TCTCCAGATA TTCCTCATGC TATTTGGGGA TTTAATGGGA CAGAACGCCC 12480

AGGAGCTGTC TATCTTGCAG CTGTACTAGC TTCACATGCT CAAAAAGGGA TTCCAGCCTT 12540

TGGGATTTAT GGAAGAGATG TTCAGGAAGC TAGTGACACA GATATTCCAG AAGATGTCAA 12600

AGAAAAACTT TTACGCTATG CGCGTGCAGC TCTTGCAACT GGCTTGATGA GAGACACTGC 12660

TTACCTATCA ATGGGTAGTG TTTCGATGGG GATTGGTGGT TCTATTGTAA ATCCGGATTT 12720

CTTCCAAGAA TACTTAGGAA TGCGAAATGA ATCGGTAGAT ATGACGGAGT TCACGCGCCG 12780

TATGGACCGT GGTATTTACG ACCCTGAAGA GTTCGAACGT GCGCTCAAAT GGGTGAAAGA 12840

AAACGTAAAA GAAGGATTCG ACCATAACCG TGAAGACCTT GTTTTAAGCC GTGAAGAAAA 12900

AGATAGACAA TGGGAATTTG TTATTAAGAT GTTCATGATT GGACGTGACT TAATGGTTGG 12960

TAACCCAAGA CTTGCTGAAC TTGGTTTTGA GGAAGAAGCG GTTGGTCACC ATGCTTTAGT 13020

AGCTGGTTTC CAAGGTCAAC GTCAGTGGAC AGACCATTTT CCAAATGGGG ACTTTATGGA 13080

AACTTTCCTC AATACTCAGT TTGACTGGAA TGGTATTCGA AAACCATTTG TATTTGCGAC 13140

AGAGAATGAT TCACTAAATG GTGTGTCTAT GCTCTTTAAT TATCTATTAA CAAATACTCC 13200

ACAAATCTTT GCTGATGTGC GTACTTATTG GAGCCCAGAG GCTGTTAAAC GTGTAACGGG 13260

ACATACTTTA GAGGGTCGTG CTGCAGCTGG CTTCTTACAT CTAATCAACT CTGGTTCTTG 13320

TACATTGGAT GGTACAGGTC AAGCTACTCG AGATGGCAAA CCTATTATGA AACCATTCTG 13380

GGAGTTGGAA GAAAGTGAAG TGCAGGCTAT GCTTGAAAAT ACAGACTTCC CACCAGCAAA 13440

CCGCGAATAC TTCCGTGGAG GAGGATTCTC AACTCGTTTC TTGACGAAGG GGGATATGCC 13500

AGTAACAATG GTACGTCTCA ATCTTCTAAA AGGGGTTGGT CCAGTGCTAC AAATTGCAGA 13560

AGGTTACACA CTTGAACTTC CTGAAGATGT TCACCATACT TTAGATAATC GTACAGATCC 13620

AGGATGGCCA ACTACTTGGT TTGCTCCACG TTTGACAGGA AAAGGTGCTT TCAAGTCTGT 13680

CTATGACGTC ATGAATAATT GGGGAGCTAA TCACGGAGCC ATAACATATG GACACATTGG 13740

AGCAGACTTG ATTACCTTGG CTTCTATGTT GAGAATTCCT GTCAATATGC ATAATGTACC 13800

TGAGGAAGAT ATCTTTAGAC CTAAAAATTG GTCCTTATTT GGAACAGAAG ATCTAGAATC 13860

AGCAGACTAT CGTGCATGTC AGTTGTTGGG GCCACTACAT AAATAAAACT TGTTTATATA 13920

GGAGGTGAAC TTACGTCCCT CCTATCCTTT TAAAAAGATT TGTTAAACAA TTCACAAATA 13980

ATTGAAAACG AATACAAAAA GTAATATAAT GATGTTAAAT AGATAGCGCG GAGGCGCAGG 14040

AGGAAAATTA TATGGCTATA TTTTATGTTC CGGCAGTCAA CCTTATTGGA AAAGGTGTTG 14100

TAAATGAAGT GGGTCCTTAT ATCAAGGAAC TTGGCTATAA AAAGGCACTT TTGGTGACAG 14160

ATAAGTACAT CGAAGGCAGT GATATTTTAC CTAAGACTTT AAAACCACTG GATACAGAAG 14220

GAATCGAATA T 14231

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16995 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

AGTTCTCTTA ACTTTTTTAG GATGGCATTC TCCGCTCTCA GGTACTCATT TTCTGCTgAA 60

GACGTTCTAA TTCTGTCCTC TCTTCAGGTC TCGTTTTTGG CTTACGTCCC ATTTTAGGTA 120

CTCTCCCTCT TGTTTTCTCA ACAATAGTAT ACCCGTTTTT CCTGTATTGT GCTAGCCAGT 180

TAAGAAGTAT CGTACGACTT GGGAGACCGT ATTCAAGAGA AACTCTATCT TTAGTCCAGC 240

CTTCATGTCA GACTTTATTA CTCATTTCTT GTTTTAAATC AGGAGAATAG TAACGATTTT 300

TTCCTTTTTT GACGAACTCT ATTCCGTAAC GATCAATCAA TTTAATCATG TACCTAATAT 360

TAGAATTGCT TATCCCAAAT TTATTTGAAA GCTTCTCTAA GCTATATCCT TGTTTTCTAA 420

GTTCATAGAT CTGAACTTTA TCATCATAAG TTAGTTTCAT AATAAAAACA CCCCAAAAGT 480

TAGATTTTTT CTGTCTAACT TTTGGGGTGT AGTTCATGTA CACCTGATAT GATGCGTTTT 540

ATAATTTTTA AGCCTTTTTG CCCAGCCTCG TCAAAAGTAA TGTTTTGACA CAAAATCTGT 600

GACAAAACTT TAGTTTTAAA GGTTTTTAAC TTTGTATATA CTAGTTTTAA GAAAAGGAGG 660

ATGATCTAAT GGAAGAAAAA GTATCATTGA AAGTCAGGGT TCAAAAACTA GGGACATCGC 720

TTTCAAATAT GGTTATGCCC AATATTGGAG CATTTATTGC TTGGGGAGTA TTGACTGCCC 780

TCTTTATCGC TGATGGCTAT CTGCCAAATG AACAGTTAGC TACTGTTGTT GGTCCTATGT 840

TAACGTATTT ATTGCCAATC CTGATTGGTT ACACAGGTGG ATATATGATC CATGGCCAAC 900

GTGGTGCCGT TGTAGGAGCT ATTGCTACTG TTGGTGCAAT CACAGGTTCT AGTGTTCCTA 960

TGTTTATCGG AGCTATGGTA ATGGGCCCAC TGGGAGGATG GACTATCAAG AAATTTGATG 1020

AGAAGTTCCA GGAAAAAATT CGTCCCGGAT TTGAAATGTT AGTTAATAAC TTCTCAGCTG 1080

GTCTCGTTGG TTTTGCATTA TTGCTTTTGG CTTTCTACGC AATCGGTCCA GTCGTATCGA 1140

CTCTTACTGG AGCTGTTGGG AATGGTGTTG AGGCTATTGT CAATGCTCGC CTCCTTCCTA 1200

TGGCTAATAT TATCATCGAA CCGGCTAAAG TCCTTTTCCT CAATAATGCC CTCAATCATG 1260

GCATTTTTAC TCCTCTGGGA GTAGAACAGG TAGCTCAAGC TGGTAAGTCA ATTCTCTTCC 1320

TATTGGAAGC TAATCCTGGA CCAGGTCTGG GAATTCTATT AGCTTATGCT GTATTCGGTA 1380

AAGGTTCTGC TAAATCTTCT TCTTGGGGGG CAATGGTTAT TCATTTCTTC GGAGGGATTC 1440

ATGAAATTTA CTTTCCTTAT GTTATGATGA AGCCTACTCT ATTTTTAGCT GCTATGGCAG 1500

GAGGTATCTC TGGAACTTTT ACTTTTCAAC TCTTAGACGC TGGTCTTAAA TCTCCAGCTT 1560

CACCAGGTTC TATTATTGCG ATTATAGCTA CGGCGCCAAA AGGTGTTTGG CCCCATCTAA 1620

ATGTTCTTTT AGGTGTTTTA GTGGCAGCAG TTGTTTCTTT CCTTGTAGCA GCCCTTATTC 1680

TTCATGCAGA CAAGTCAACT GAGGATTCGC TCGAAGCTGC TCAGGCGGCT ACCCAAGCAG 1740

CTAAGGCTCA GTCTAAAGGT CAGTTAGTAT CAACTTCTGT TGATGCAGTT GTTTCGACAG 1800

ACTCAGTGGA AAAAATCATT TTCGCCTGCG ATGCTGGTAT GGGAAGCTCT GCTATGGGAG 1860

CTAGTATTCT TCGAGATAAG GTTAAAAAAG CAGGTCTAGA GATTCCAGTA TCTAATCAGG 1920

CAATCTCAAA TTTGCTTGAT ACACCAAAAA CATTAATTGT TACTCAGGAA GAACTGACAC 1980

CAAGAGCTAA AGACAAGAGT CCAAGTGCTA TTCATGTTTC TGTTGATAAT TTCTTAGCGT 2040

CCTCTCGTTA TGATGAAATT GTAGCTTCAT TAACAGGAGC TTCTCCAATA GCAGAAATTG 2100

AAGGAGATAT ACCAACTTCA GCACCAGTAG ATAGTCAGGA AAGTGACCTT AACCATATTG 2160

ATGCTGTAGT AGTTGCTTAT GGTAAAGCAC AGGGAACTGC AACTATGGGC TGTGAAACGA 2220

TTCGGGCTAT TTTTAGAAAC AAGAATATTC GTATTCCAGT TTCTACTGCC AAAATTTCAG 2280

AATTAGGTGA ATTTAATTCT AAAAACATAA TGATTGTAAC AACTATTTCT TTACAGGCAG 2340

AAGTGCAGCA AGCAGCACCG AATTCTCAAT TTCTTATTGT GGATAGTTTA GTAACAACAC 2400

CAGAATATGA CAAAATGGCT GCTAGAATGT ACAAATAGAA CTAGAGGTTT CTAAATTACG 2460

AATGCTATTA ACCAAACGAG AAGAACAATT ATTGAAGGCT TTCCTACATG TAGGGAAGCT 2520

TTCAATGCAA GATATGACTG AAATCTTACA GGTTTCATCT AGAACAATTT ATCGAACTTT 2580

ATCAGATTTG ACAGATAGCA TGGAGCAATA TGGAATCGAA ATAACGAAGC ATGGGAAATA 2640

CTATATTTTG ACTGGAGAGT TGGATGATTT GCCGACAGAA CTTGAAGTGT TAGTTGAGTA 2700

TAGTCCCCAA GAAAGACAAG AGTTGATTAC CTATCGCCTT CTGACTGAGA GTGGTTTTGT 2760

CACCAATGAA GCATTGCAAG AGTGCACGAA AGTCAGTAAT GTAACTATTA TTCAGGATAT 2820

TTCAGATATT GATAAGCGTC TTTTAGACTT TGATCTGAAA ATTGAACGAC AAAAAGGTTA 2880

TCGGATTTCT GGTGATTCAG TTGGTAAGAG AAGATTTTTG GCTATTTTAC TGACAAACTG 2940

TATCTCAGTA GCAGATTTTT CAACCGGTAA TTTTGGGAGC TTTGATATTT TAGAAGCAGA 3000

TAGAACTGGG CTGGCCAGTC AGATTGTTAA TAAGCAACTG TCAGGTTTTC CAGATATGGA 3060

TGCTAGGATG AAGATGTTTT TTGCGATCTT GTTATCTCTT ATAGGTCAGG AGCAAAACAT 3120

TGAAAATTCA CCTAATACTA GTAAGCAGGC TTTGGAAATT TCTCAAAAAA TTTTTCAAGC 3180

TTACTCTAAG CAGACTGCAC AATTTTATAG TATTCAGGAA ATTATCTATT TTGCGAGCAT 3240

CTTGGATGAA TTAATCATTA AACGTCAGGA CAATCCGCTC TTTACGGAGA AATTTGATGG 3300

TGAATTTTTC TACAATATTT CAAATCTGAT TGATACGGTT TCCATGTATA CCAAGATTGA 3360

CTTTTTTAAG GACAAGGTTT TATTCAATTT TCTTTTCCAT CATATTCGGC TCAGTTTAGG 3420

CGTCCCTATC CTTTTTCAGG GTGAAAATTT GCCAGAATCT ATCCAGATTT TAGTTGAAAG 3480

GAATAAATTT CTTTATACAG TCATCAGTCT TTTAGTGAAT GATATTTTTC CGAAATATCT 3540

TCATACAGAG TATGAGTATG GCATGATTGC CCTACATTTT ATCTCTAGCT TAGGCCGTAG 3600

TCCAGAGATT TATCCAGTCC GTGTTTTGCT TTTAACGGAT GAACGTCGGG TCACTAGAGA 3660

TTTATTAGTC AGTAAAATTA AGAGTGTTGC TCCTTTTGTA GAGTTGATAG ATATTCAATC 3720

TCTAGTAGAT TACCACAGTA TTGATCTCAG TCAGTATGAT TATATTTTAT CTACCAAGCC 3780

GCTGACTAAT CAGGAAATCG ATGTAATTTC TAGTTTTCCA ACCGTCAAAG AATTGCTTGA 3840

ATTACAGGAA CGACTTCAGT ATGTACAGGC ACATCGTACA ATTGTCGCGC GTGATGCTAT 3900

CGCTCCAGAG AAAAGTTATG ACTTGCAAGA TTATTTAATA TCTAGTAGTC AGCTTTTGAG 3960

TCAATTCGAG TTGGTTCAAT TGGAGAATAA TCAATCATTT GAGCACACGG TAGAACAAAT 4020

CATCCAATAT CAGAAGAATG TGAGTGACAG AGCTTACCTA ACAAGAAAAT TGTTATCTCA 4080

CTTCCAGAAT AGTCCTATGG CTATTCCTAA TACTGGTCTG GTGCTTTTAC ATAGTCAGTC 4140

TAGCAAAGTA ACAACAAATA GTTTTACTAT GTTTGAACTC AAACTACCTA TCTCCGCATT 4200

GTCAATGAAA CGAGAGGAAG AAGAGGTCAA AAGGTGTCTG CTAATGCTAA TGTCTAAAGA 4260

AGCTAGCGAG GAAGCTAGAG ATTTAATGAC AGCTATTAGT CAGTCGATTA TTGAAAATCA 4320

TCTTTATACA GAGATTTACA AGACGGGAAA TCAATCCATT ATTTATCAGA TGCTAAATAC 4380

TATTTTTAAC GAAAAAATTA AGAAATTGGA GAACTAATAT GAAACTTGAA AAACATTTGA 4440

TTAAGCTTAA TAAACAATTT TCTAACAAGG AGGAAGCTAT TTGTTATTGT GGGCAAGTTC 4500

TTTATGAGGG TGGATATGTT AATGAAGACT ATATTGAAGC CATGATTGAG CGAGATAAAG 4560

AGCTATCTGT TTACATGGGT AACTTTATCG CCATACCGCA TGGAACAGAT GCAGCAAAAA 4620

ATGATGTCCT CAAGTCTGGT ATTACAGTCG TTCAAGTCCC TAGAGGGGTT GATTTTGGGA 4680

ATGTATCTAA CCCTCAAGTG GCAACGGTTC TTTTTGGTAT TGCTGGTATT GGTAATGAAC 4740

ACTTAGAAAT TATTCAGAAA ATTTCTATCT TCTGTGCAGA TGTAGATAAT GTTCTTAAAC 4800

TAGCAGATGC TCAGTCAAAA GAGGAAGTAT TGCGCTTATT TGATGCTGTT GAATAATTGA 4860

ATTTAGTCAT TTGTCATCTA GTATATATGT CCCTCAAATA GGAAAAGGAG AAATTGAATG 4920

AAACATTCTG TTCATTTTGG TGCCGGTAAT ATCGGTCGTG GTTTTATAGG TGAAATTCTA 4980

TTTAAAAATG GTTTCCATAT TGATTTTGTG GATGTCAATA ATCAGATAAT TCATGCTCTG 5040

AATGAAAAGG GCAAGTATGA AATTGAAATT GCACAGAAAG GACAGTCTCG TATAGAAGTA 5100

ACTAATGTGG CTGGCATTAA TAGCAAAGAA CATCCTGAGC AAGTCATTGA AGCGATTCAA 5160

AAGACGGATA TTATTACTAC TGCAATCGGA CCTAATATAC TCCCTTTTAT CGCCGAACTT 5220

CTAGCCAAAG GAATCGAAGC TCGCCGAGTT GCAGGAAATA CACAGGCATT GGATGTTATG 5280

GCCTGTGAAA ATATGATTGG CGGGTCTCAA TTTCTTTATC AAGAAGTCAA GAAATATTTA 5340

AGTCCGGAAG GTTTGACATT TGCTGATAAC TACATAGGTT TTCCAAATGC TGCAGTAGAC 5400

AGGATTGTTC CAGCACAAAG TCACGAAGAT TCCCTTTTTG TTGTGGTCGA GCCCTTTAAT 5460

GAATGGGTCG TGGAAACCAA GCGTCTTAAA AATCCAGATT TACGTCTAAA AGATGTGCAT 5520

TATGAAGAAG ATTTAGAACC CTTTATTGAG CGAAAACTTT TTTCAGTCAA TTCTGGACAT 5580

GCAACTTCAG CTTACATTGG TGCGCATTAT GGTGCCAAGA CAATTTTGGA AGCTCTTCAA 5640

AATCCTAATA TTAAATCTCG GATTGAATCT GTATTAGCTG AAATTCGGAG TCTCTTGATT 5700

GCCAAATGGA ACTTTGATAA AAAAGAATTG GAGAATTATC ACAAAGTCAT TATAGAACGA 5760

CTTGAAAACC CTTTCATAGT GGACGAGGTT AGTCGCGTAG CTCGTACTCC AATCCGAAAA 5820

TTAGGCTATA ATGAACGATT CATCCGGCCG ATACGTGAAT TGAAAGAACT CAGTTTGTCA 5880

TATAAAAACC TACTTAAAAC AGTTGGCTAT GTCTTTGACT ATCGCGATGT AAATGATGAA 5940

GAAAGTATTC GATTAGGTGA ATTGTTGGCT AAACAATCAG TCAAAGATGT TGTTATACAA 6000

GTTACAGGTT TAGACGACCA AGAATTGATT GAGCAAATTG TAGAGTATAT TTAATCTTTT 6060

TCGAAAATCT CTTCAAATCA GGTTAGCATC GCTTTGTCTT AGGCATATGT TGTTCTATCT 6120

ACAACCTCAA AGCAGTGCTT TGAGCTGACT CCGTCAGTCT TATCTGCAAT CTCAAAACAC 6180

TGTTTGAGTT ATCTGCGGTA ATCTTTCTAG CTTGTCTTTG ATTTTTGTTG TTATTTATAA 6240

GGTAAAAGAA GCTGGACAAA AAGTCTTCAA AATCGGGAAA AGGCAGCCTA TCGGGTGTTC 6300

AAAAATCTTG ATAGGATGTC CTTTATTATG GAAAGCCTTA TTGGATTTTC TCCTCAGATT 6360

GAGTTTTTGA TCAGCTTTAT GAGATAGGTC TTGCTAGAGA TGTAGCCCAT CATGTTATTT 6420

TTATGGACAG TGGGAAAATT GTTGAAAAAA ATAATGCCCA TCAATTCTTT AGTCGTCCAA 6480

GAGAAGAACG AACCAAGCAA TTTTGGAACG AATTCTTTCG AATGCGATCT ATATAGTAAA 6540

ATGAAACAAG AACAGGACAA ATCGATCAGG ACAGTCAAAT CGATTTCTAA AAATGTTTTA 6600

GAAGTAGAGG TGTACTATTC TAGTTTCAAT CTACTATATA ACTGAAAAAT TAGATAAATT 6660

AGTTTTGGAA AATGACTAAC CAAAAGATAT CCAAAGTAGT CTAAAATTGT CTATACTTTA 6720

TGAGTGTTTT AGTTAGGAAA AAGGCTTGTT GTCTATAATT GTCTGCATTA GTCTAGATTT 6780

TATTTATAGA AAATGTTATA ATAGACTGTA TTTAAAAAAT TTTAAGGAGA AATGACAGAA 6840

TGTCTGTATC ATTTGAAAAC AAAGAAACAA ACCGTGGTGT CTTGACTTTC ACTATCTCTC 6900

AAGACCAAAT CAAACCAGAA TTGGACCGTG TCTTCAAGTC AGTGAAGAAA TCTCTTAATG 6960

TTCCAGGTTT CCGTAAAGGT CACCTTCCAC GCCCTATCTT CGACCAAAAA TTTGGTGAAG 7020

AAGCTCTTTA TCAAGATGCA ATGAACGCAC TTTTGCCAAA CGCTTATGAA GCAGCTGTAA 7080

AAGAAGCTGG TCTTGAAGTG GTTGCCCAAC CAAAAATTGA CGTAACTTCA ATGGAAAAAG 7140

GTCAAGACTG GGTTATCACT GCTGAAGTCG TTACAAAACC TGAAGTAAAA TTGGGTGACT 7200

ACAAAAACCT TGAAGTATCA GTTGATGTAG AAAAAGAAGT AACTGACGCT GATGTCGAAG 7260

AGCGTATCGA ACGCGAACGC AACAACCTGG CTGAATTGGT TATCAAGGAA GCTGCTGCTG 7320

AAAACGGCGA CACTGTTGTG ATCGACTTCG TTGGTTCTAT CGACGGTGTT GAATTTGACG 7380

GTGGAAAAGG TGAAAACTTC TCACTTGGAC TTGGTTCAGG TCAATTCATC CCTGGTTTCG 7440

AAGACCAATT GGTAGGTCAC TCAGCTGGCG AAACCGTTGA TGTTATCGTA ACATTCCCAG 7500

AAGACTACCA AGCAGAAGAC CTTGCAGGTA AAGAAGCTAA ATTCGTGACA ACTATCCACG 7560

AAGTAAAAGC TAAAGAAGTT CCGGCTCTTG ACGATGAACT TGCAAAAGAC ATTGATGAAG 7620

AAGTTGAAAC ACTTGCTGAC TTGAAAGAAA AATACAGCAA AGAATTGGCT GCTGCTAAAG 7680

AAGAAGCTTA CAAAGATGCA GTTGAAGGTG CAGCAATTGA TACAGCTGTA GAAAATGCTG 7740

AAATCGTAGA ACTTCCAGAA GAAATGATCC ATGAAGAAGT TCACCGTTCA GTAAATGAAT 7800

TCCTTGGGAA TTTGCAACGT CAAGGGATCA ACCCTGACAT GTACTTCCAA ATCACTGGAA 7860

CTACTCAAGA AGACCTTCAC AACCAATACC AAGCAGAAGC TGAGTCACGT ACTAAGACTA 7920

ACCTTGTTAT CGAAGCAGTT GCCAAAGCTG AAGGATTTGA TGCTTCAGAA GAAGAAATCC 7980

AAAAAGAAGT TGAGCAATTG GCAGCAGACT ACAACATGGA AGTTGCACAA GTTCAAAACT 8040

TGCTTTCAGC TGACATGTTG AAACATGATA TCACTATCAA AAAAGCTGTT GAATTGATCA 8100

CAAGCACAGC AACAGTAAAA TAATCTTAAT AAACAGAAAA CCCACCTGAA TTGGTGGGTT 8160

TTCTGATGCA CTATTTTCCA AAAATCTCTT TGAGGTCTGT GTCTGTAATC CCAATCATGG 8220

CTGGGATGCG GTCCCAGTTT TCTTCGGTTA GGATGTAGGA TTGTTCAGAG GCACTTGATG 8280

TGACTGTTTC AGAGACAGCT TGTTGCTTTT CTTCAACATT CTCCAGTAGA TCACTGAAGC 8340

GTTCAATCAG ATAGGTTTTT CGGGCAGTTC CGATGTGTTG GGTAGCATAG TCGAAGGCTT 8400

GTAATTCGCC TAGTAAGATG AGTTTGCTTT TGGCACGTGT AATGGCTGTG TAGATGAGAT 8460

TTCGCTCCAG CATACGTCGG CTAGCACTAG TAATCGGTAG GATGACAACT GGGAACTCAC 8520

TTCCCTGAGA CTTATGAATA CTCATGGCAT AGGCCAAGCG AATCTTGTAC CATTCGTTAC 8580

GGGGGTAAGA GACTTCATTA CCATCAAAAT CAATGACAAT CTCGTCTTGT TTCGATTCGG 8640

TGTATTTACC AGGAATCAGG TCTGTGATAG CTCCTAAATC CCCATTAAAG ACATTGATTT 8700

CAGCATCGTT AACCAAATGA ATGACCCTGT CTCTCTTACG ATAGTGACAC TGAGGAGCTT 8760

CAAAACTGAG TTGATCTTTT TGTGGGGGAT TGAGCAGGTC TTGCATGAGC TGATTGATAG 8820

CATCAATCCC TGCCGTCCCT CGGTACATAG GAGCCAGAAC TTGGATATCA CGGGCGGGAA 8880

TACCATTTCT GAGGGCGGCA CCTAAGATTT TTTCAATGGT GGCAGGAATA TGGCCACTAG 8940

CAATTTCAAA GTAGGAACGG TCAGCTTTTT TTTGGGTGAA ATCAGCTGGC AAGATGCCCT 9000

GTCGAATCTG ACTAGCTAGG GTGACGATGG TTGATTCTTT GCTTTGTCGA TAAATTTTTT 9060

CCAAGCGAGT CTGAGGAATC AAAGGAATAT GAAGTAGATC CGCTAGAACC TGTCCAGGAC 9120

TGACAGAAGG TAGCTGATCA CTGTCACCTA CGATGAGGAT CTTACTGTTA GAAGAGATAT 9180

TGGAGAAGAG TTGATTGGCC AGCCAAGTAT CTACCATAGA GAATTCATCC ACGATGATAA 9240

AGTCAGCATC TAGGTAATCT TCCAGATGAC TGGTATCATC GTCACCTGTC ATTCCCAAGT 9300

GGCGATGTAT GGTCGCGCTA GGCAAACCTG TCAATTCATT CATGCGACGA GCAGCTCGAC 9360

CAGTTGGAGC AGCAAGAAGA ATGGGCAGAT TGCTTTTCTT CCTGAAGTCA AGTCCTTCTA 9420

AAAGGGCATA AACAGCAATG ATTCCATTGA TAACAGTTGT CTTACCAGTA CCAGGCCCAC 9480

CTGTCAGGAT AAAGACCTTA TTCTGGATAG CATCACAGAT AGCCTGTTTT TGAATGTTAT 9540

CATACTCAAT TCCCAGTTCT TGCTCGACAG TAGTGATATG TTTTTGAATG GTTTCTAAAT 9600

CATGACTCTT CTGTTTTCCT TTTTCAAGGA TACGAACCAA GTGACTGCGG ATGCCTTCCT 9660

CAGCGAAAAA GAGGCTGTTG TCAAAGATCT TGGTATCAAT CTGCTGAACC TTGTCTTCTT 9720

CGATCAGGTA GGAGAGCTCT TGGGCAACTT GGCTGGGGTC TAGTTCCACG GGACGGGAAG 9780

ACTCAAGGAG AGTAAGGGTT TGTTCCAGCA AATCCCGTGC TTCAACATAG GTGTCCCCTG 9840

TTTCCATACA GGCCTGAAAA AGACTGTGAA CTAGACCGGC GCGGAAGCGT TCAGGAGCCT 9900

GACTTTCGAT GCCTAGTTCC TCAGCTAGTT GGTCAGCAAT GGTAAAGCCC AAACCCTTGA 9960

TATCCTCAAC CAGCTGGTAG GGATAATTTT CAACCACATC AAGGGTTTCT TCCTTGTAAA 10020

AGTCTTGAAT CTGAAAGGCT AGTTTGTTGG GAATGCCGTA GTTGGCTAGT TTGGCCAAAA 10080

TCATCTCCGT TCCGTAGTTG AGACGGAGAG TGGAGACGAA AGCCTCGCGA TTTTTGGCAG 10140

AGAGTCCTGC GATGCCTTCT AACTTTTCTG GGTGTTGCAA AATTTCGTCA ATGGTATTTT 10200

CGCCATAGGT ATCCACGATT TTCTGAGCTG TCTTGAGACC AATCCCCTTG AAATGGCTAC 10260

TTGAAAAGTA CTTGACCAAG CCCTTACTAG TTGGTTTTGC GCGATCATAA CGACTGATTT 10320

GCAGTTGTTC TCCATACTTG GAGTGCTGGA CAATTTGCCC CCAAAAAGTA TAGTCTTCGC 10380

CCTCAATTAC ATCAGCCATG GTTCCTGTGA CAATGATTTC AAAATCATCA AAATCCTCTG 10440

CGTCCGTATC GTCGATTTCT AGGAGGAGGA TGCGATAAAA ATTGCTGGGA TTTTCAAAAA 10500

TAATCCGTTC AATAGTTCCT GAAAAATAAA CTTCCATAAA ATTCCTTTGC ATGAATAGGT 10560

GAGAGTTGGG ATTGTTTTTA TTTTATACTC TTCGAAAATA TCTTCAAACC ACGTCAGCTT 10620

CCATCTGCAA CCTCAAAACA GTATTTTGAG CTGACTTCGT CAGTTCTATC CACAACCTCA 10680

AAACACTGTT TTAAGCAGCC TACGGCTAGC TTCCTAGTTT GTTCTTTGAT TTTCATTGAG 10740

TATTTGTAAA TAAACAATCA CTTCTCACGA TAGAAGAAGA GGCTGAGATT GGTGATTCTC 10800

TGCCTCTTAG GTTTCTTAAA ATGTTCCGAT ACGGGTGATT GGCCATAAGC GGAATTTAGC 10860

TTCCCCTGTG ATATCTTTTG CTTTGAAGGT ACCTACGTGG CGGCTGTCGC TCGAAACCAA 10920

GCGGTCATCT CCGAGGAGAA GGTATTCTCC TTCTGGAACA GTAAAGCTAA AGTTGGTGTT 10980

GTAGTTGACA TCAACTGTGA AGGCTTGAGC TTTTTGAGCG ATACTTCTAA AGAAAGTTCC 11040

TTTATTTCCT TCAAAGCCCT TGCCTGAGTA AGTGCTTTGG AGTTTGTCAT CCTTGAAGCG 11100

TTTGATATAG TCTGCTAGAT AAGGCTCGTC CGTTTCTTTG TCATTGATGT AGAGTTTATC 11160

ATTTTCGTAA CGAATGGTGT CGCCAGGCAT TCCAATCACG CGCTTGACGA TGTCCTTATT 11220

GCCATCTTCC TCATGGGCCA CCACGATATC AAAACGGTCA ATAGGAAGGT GTTTTACAAC 11280

GAAGAGAATT TCGCCATCCG CTAGGGTCGG ATCCATGGAA TGTCCTTCTA CGCGAACATT 11340

GCTCCAAAAA AAGATACGAC TTAAAGCTAG TAATGACAGA ATTAGGAGGA ACAATCCCCA 11400

CTCTTTTAAG AAATTTTTAA ATGAATTCAT AACTTACCTT TCTAAGCGTT TTTTCGCTTT 11460

TTCAGTGTTT TTAAAGTGCA ATTTGGCGCA GAAGCTGAGT CCCTGCATAC CATAGGCTTG 11520

CAAAATCTGG CTAGCCACCT TGTCAGAAGC CGTTCCAGCT CCACTTGGGA GCTGATAACC 11580

CAGTTCTCGT CCCAAATTTT CAAGATTTTC CAGAAAGAGA TCACGCGCAA TGACAGAAGA 11640

AACTGCGACA GACAAGTATT TGCCCTCAGC CTTTTCTTCT AAGCTGATAG GATTGCTGAA 11700

ACGATTGGCC TCTTGTGCCA AGTACTTGTC ATAATTTTTA GCACTGGTAA AGGCATCAAT 11760

CACAATTTTC TCAGGCTGAA CACCTTTTTG AAGGAGGAGA TAGATAGCCT GATTATGGAG 11820

GGCAACCTTA ACCGAAACAG CGTTGTAGCG GTCTCCGATG ACCTCGTTGT ACTTGCTGGG 11880

TGAGAGAAGG AGTGCCTGGT GCTGAATTTT TTCCTTGAGA ATAGGAGTAA TCTGACGGAT 11940

CTTTTGGTCG GTCAGAGTCT TAGAATCCCC CACACCGAGT TTTCGTAAAA AGTCGTGCTG 12000

GTCAGGTGTG ACAAAGGCAG CCACAACTGC AAGCCCACCA AAGTAGGAAC CATTTCCCAC 12060

CTCATCTGTC CCAATTAAAG GAAGATTTTG TCCGCTGGTT TGCTCTACAG CTTGATAGCC 12120

AAAGAAACTG GCGTATTTTT CAGCCCCTTC ACCCTGAAGC AAGATTTTTC CAGAAGTATA 12180

GATAGAAACC GTTGCTTGAG GTAGTTTCAA AAAGTAGCGG ATATAGGGAT TCTTGCTAGG 12240

AGCCAGACTG GTTTGATAGT GTTCAAGAAA AGCCTGAATA TCCTTTTCGC TTGGTGTGAG 12300

TGTGATACTT GCCATAGTTT CTATTGTACC ACAAAAGCAG TAAAATTTGT AAAAACTGAC 12360

AAAATTAGCG AATTTTGGTA TAATATCGTG AGGTGAATTT TATGGCAAAT CTAAATCGAT 12420

TCAAATTTAC ATTCGGGAAA AAATCGTTAA CCTTGACAAG CGAACATGAC AACCTTTTTA 12480

TGGAGGAAAT CGCTAAGGTT GCGACAGAAA AATACCAAGC AATTAAAGAA CAAATGCCTA 12540

GCGCAGATGA TGAAACAATC GCTCTTTTGT TGGCAGTCAA CTGTTTATCA ACTCAGCTCA 12600

GCCGTGAGAT TGAATTTGAC GATAAGGAGC AAGAGCTAGA AGAACTCCGT CACAAGCTTG 12660

TGACTTGTAA GCAAGAACAG AGCAAGATTG AGGATTCCTT ATGATTTCAT TCCTTCTTCT 12720

ATTGGTCTTG GTTTGGGGAT TTTATATCGG CTATCGGAGA GGCCTGCTCT TACAGGTTTA 12780

TTACCTGATT TCAGCCATGG CATCGGCTTT TATGGCTGGC CAGTTTTATA AGGGGCTTGG 12840

AGAGCAATTC CATTTATTGC TCCCTTATGC AAATTCGCAG GAAGGTCAGG GGACTTTCTT 12900

TTTCCCATCG GATCAACTCT TTCAGCTGGA TAAGGTCTTT TATGCAGGTA TCGGCTACTT 12960

GCTTGTATTT GGGATTGTCT ATAGCATTGG TCGTTTACTT GGTCTTCTCT TACACTTGAT 13020

TCCTAGCAAA AAACTGGGTG GTAAGTTGTT CCAAGTTTCA GCAGGTATCT TGTCCATGTT 13080

GGTGACCTTA TTTGTCTTGC AAATGGCCTT GACAATCTTG GCGACCATCC CCATGGCAGT 13140

TATACAAAAT CCTCTTGAAA AGAGTATCGT CGCAAAACAC ATCATCCAGA GCATACCGGT 13200

AACAACCAGT TGGCTCAAAC AAATCTGGGT GACAAATTTA ATCGGATAAA AAGGGCAGGA 13260

GTTTTCCTAG CCCTTTGTTT ACAGATTTGA CTCGAATCTA TCAGAATGTA AAAAGCTACC 13320

ACACCTAGAC ATTCAAAGAC AAGGAAATAA AGATGAATAA GAAAATATTA GAAACATTAG 13380

AGTTCGATAA GGTCAAGGCC TTGTTTGAGC CTCATTTGTT GACCGAGCAG GGCTTGGAGC 13440

AATTGAGACA ACTGGCTCCG ACTGCCAAAG CAGATAAAAT CAAACAGGCT TTTGCTGAGA 13500

TGAAGGAAAT GCAGGCTCTT TTCGTCGAGC AACCGCATTT TACTATTCTC TCAACTAAGG 13560

AAATTGCAGG AGTCTGCAAG AGGTTGGAGA TGGGAGCGGA TCTCAATATC GAGGAGTTCC 13620

TACTCTTGAA ACGCGTGCTT CTTGCCAGCC GAGAACTTCA AAATTTTTAC ACCAATCTGG 13680

AAAATGTCAG CTTGGAAGAA TTAGCCCTTT GGTTTGAGAA ATTACATGAT TTTCCGCAAT 13740

TACAAGGAAA TCTTCAGGCC TTTAATGATG CGGGTTTCAT TGAAAATTTT GCCAGTGAAG 13800

AATTGGCGCG AATCCGTCGA AAAATACATG ATAGCGAGAG TCAGGTACGC GATGTTTTAC 13860

AAGACTTGCT CAAGCAAAAA GCGCAGCTGT TGACGGAAGG AATTGTTGCT AGCAGAAATG 13920

GCCGTCAGGT TTTACCAGTC AAAAACACCT ACCGCAATAA GATTGCAGGT GTCGTTCATG 13980

ATATTTCTGC TAGTGGAAAC ACCGTCTATA TCGAACCCCG TGAGGTAGTC AAACTGAGCG 14040

AAGAAATTGC TAGTCTGCGA GCAGATGAGC GCTATGAAAT GCTTCGCATT CTCCAAGAAA 14100

TTTCTGAGCG TGTCCGCCCT CATGCGGCTG AGATTGCTAA TGACGCTTGG ATTATCGGTC 14160

ATCTGGACTT GATTCGTGCC AAGGTTCGAT TTATCCAAGA AAGACAAGCA GTCGTGCCTC 14220

AGCTGTCAGA AAATCAAGAG ATTCAACTGC TCCATGTCTG CCATCCTTTG GTCAAAAATG 14280

CCGTCGCAAA TGATGTCTAT TTTGGTCAAG ATTTAACAGC TATTGTCATT ACAGGTCCCA 14340

ATACAGGTGG GAAGACCATC ATGCTCAAAA CTCTGGGCTT GACACAGGTC ATGGCCCAGT 14400

CAGGATTGCC GATTTTAGCA GACAAGGGAA GTCGTGTTGG TATTTTTGAA GAAATCTTTG 14460

CTGATATTGG AGATGAGCAG TCTATTGAGC AGAGCTTGTC TACCTTCTCT AGTCATATGA 14520

CCAATATCGT GGATATTCTT GGCAAGGTCA ACCAACATTC ACTCTTACTT TTGGATGAGT 14580

TGGGGGCTGG TACTGATCCC CAAGAGGGAG CAGCCCTTGC CATGGCTATT CTGGAGGACC 14640

TTCGCCTGCG TCAAATCAAG ACCATGGCGA CGACCCACTA TCCAGAACTC AAGGCCTACG 14700

GTATTGAGAC AGCCTTTGTG CAAAATGCCA GTATGGAGTT TGATACTGCA ACTCTTCGCC 14760

CGACCTATCG CTTTATGCAG GGTGTTCCTG GCCGAAGTAA TGCCTTTGAA ATTGCCAAAC 14820

GTCTAGGCCT ATCTGAAGTT ATCGTAGGAG ATGCCAGTCA GCAGATCGAT CAGGACAATG 14880

ACGTCAATCG TATCATTGAG CAATTAGAAG AGCAGACGCT GGAAAGCCGC AAACGTTTGG 14940

ACAATATCCG TGAGGTGGAG CAAGAAAATC TCAAGATGAA CCGTGCGCTA AAAAAACTCT 15000

ACAACGAGCT TAATCGTGAA AAGGAAACCG AGCTTAACAA GGCGCGTGAA CAGGCTGCTG 15060

AGATTGTGGA TATGGCCCTA AGTGAAAGTG ACCAGATTCT CAAAAATCTC CACAGTAAAT 15120

CCCAACTCAA GCCCCACGAA ATCATTGAAG CCAAGGCCAA GTTGAAAAAA TTGGCTCCTG 15180

AAAAAGTGGA CTTGTCTAAA AATAAGGTCC TTCAAAAGGC CAAGAAAAAA CGAGCTCCAA 15240

AGGTGGGAGA TGATATCGTG GTTCTCAGTT ATGGTCAGCG TGGTACCTTG ACCAGTCAAC 15300

TCAAGGACGG TCGCTGGGAA GCCCAAGTTG GCTTGATTAA GATGACCTTG GAAGAGAAAG 15360

AGTTTGATCT TGTTCAAGCC CAGCAAGAAA AACCAGTCAA GAAGAAACAG GTCAATGTTG 15420

TGAAACGAAC TTCTGGGCGA GGACCTCAAG CTAGACTGGA TCTTCGAGGC AAGCGCTATG 15480

AAGAAGCCAT GAATGAGCTA GATACCTTCA TCGACCAAGC CTTGCTTAAC AATATGGCTC 15540

AAGTTGATAT CATCCATGGT ATCGGAACAG GAGTCATCCG TGAAGGAGTT ACCAAATACT 15600

TGCAAAGAAA CAAACATGTC AAGAGTTTCG GCTATGCCCC ACAAAATGCT GGAGGCAGTG 15660

GTGCGACTAT TGTCACTTTT AAAGGATAGC AGTATTCTGG ACTTTATAAA GTAAAAACTG 15720

TTGAACTAAT TTTTACTAAT AAACACATTG ACAAAAGCCA ACATTTTTTG TAAAATTAGA 15780

ATCAATTAAA TACCAACACC GAATGAAGTT TAATAGAAGT GGGGAATCGT TTGATTTTCC 15840

ATGACTGTAA ATGGACGGAA CTCTGGAGAG ACCGTAAAGG CACCGAAGGG CAAGGCAGGC 15900

AACTGCTCAA ACTCTCAGGT AAAAGGACAG AGCTAGGATA GACCGCTTTT TAGCATTTAT 15960

CTAAGCATTC CAGAGTACAT GTATCTTGCA TGTGCTCTTT CTTTTGGGGT TGAAACGATA 16020

GGAGAAGGAA ATGTTAGAAT TGCTTAAATC AATCGATGCT TTTGCTTGGG GACCGCCCCT 16080

CTTGATTTTA TTGGTCGGAA CAGGGATTTA CCTAACTATT CGGCTAGGAC TCTTGCAGGT 16140

TTTGCGTCTA CCCAAGGCCT TTCAGCTTAT TTTTATCCAG GATAAGGGAC ATGGTGATGT 16200

ATCCAGTTTT GCAGCTCTGT GTACAGCCTT GGCATCAACT GTTGGAACAG GAAATATCAT 16260

AGGAGTTGCG ACGGCTATCA AGGTTGGTGG ACCAGGAGCT CTATTTTGGA TGTGGATGGC 16320

GGCTTTCTTT GGAATGGCTA CCAAGTATGC GGAAGGACTC TTGGCCATCA AATACCGCAC 16380

CAAGGACGAC CATGGTGCAG TAGCGGGAGG TCCCATGCAT TATATCCTTC TAGGGATGGG 16440

AGAAAAGTGG CGACCACTTG CTGTTTTGTT TGCAGTAGCA GGAGTATTGG TTGCTCTCTT 16500

GGGAATCGGA ACCTTCACCC AAGTCAACTC GATTGCAGAA TCTATCCAAA ATACAACGAC 16560

GATTTCGCCA GCCATCACAG CTCTCGTCTT GTCTGTCTTT GTAGCGATTG CAGTCTTTGG 16620

TGGACTCAAG TCTATTTCTA AGGTTTCAAC TACTGTTGTT CCTTTTATGG CCATCATTTA 16680

TATCTTAGGA ACTCTTACAG TTATTTTCTT TAATATCGGA AAAATCCCTG GCACAATCGC 16740

TTTAGTCTTT ACCTCAGCTT TTAGTCCCCT TGCTGCGGTA GGTGGATTTG CTGGTGCTAG 16800

CGTTCGGATG GCTATTCAAA ATGGTGTGGC GCGTGGTGTG TTCTCAAACG AATCTGGTCT 16860

GGGTTCTGCT CCTATTGCAG CTGCAGCTGC CAAGACAAAT GAACCAGTAG AGCAAGGTTT 16920

GATTTCCATG ACAGGAACCT TTATTGATAC CCTCATCATT TGTACTCTAA CTGGTTTGAC 16980

CATCTTGGTA ACTGG 16995

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28473 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

CCGGGGCTTT TGTAGTATAA TAGAGATACG TTTTGAAAGT AGGAGGTATC TATGGACTTA 60

ACTAAGCGCT TTAATAAACA GTTAGATAAA ATTCAAGTTT CGTTGATTCG TCAGTTTGAC 120

CAGGCTATTT CGGAGATTCC TGGGGTCTTG CGTTTGACCT TGGGGGAACC TGATTTTACA 180

ACGCCAGACC ATGTCAAGGA GGCGGGCAAG CGAGCGATTG ATCAGAACCA ATCCTACTAT 240

ACAGGGATGA GTGGTCTGCT GACTCTACGT CAGGCAGCCA GTGACTTTGT TAAGGAAAAG 300

TACCAACTGG ACTATGCTCC TGAAAATGAA ATCTTGGTTA CAATTGGGGC GACAGAGGCT 360

TTATCTGCGA CTTTGACGGC TATTTTGGAA GAGGGAGACA AGGTACTTTT GCCAGCTCCT 420

GCTTATCCAG GCTATGAACC GATTGTTAAC TTAGTTGGGG CAGAAATTGT TGAGATTGAT 480

ACGACTGAAA ATGGTTTTGT CTTGACTCCT GAGATGTTGG AGAAGGCCAT TTTGGAGCAG 540

GGTGATAAGC TCAAGGCGGT TATTCTCAAC TATCCAGCCA ATCCGACAGG AATTACCTAC 600

AGTCGAGAGC AGTTAGAGGC CTTGGCAGCT GTTTTACGCA AGTACGAAAT TTTTGTTGTC 660

TGTGATGAGG TTTACTCAGA ATTGACCTAC ACAGGCGAAG CCATGTGTCT CTAGGAACGA 720

TGTTGAGAGA CCAGGCTATT ATTATCAATG GTTTGTCTAA ATCGCATGCC ATGACAGGTT 780

GGCGTTTGGG GCTGATTTTC GCTCCTGCGA CCTTCACAGC CCAGTTAATC AAGAGTCACC 840

AGTACTTGGT CACTGCCGCA AATACCATGG CGCAACATGC TGCGGTAGAA GCCTTGACGG 900

CTGGTAAAAA CGATGCGGAC CCATGAAGAA GGAATATATC CAACGTCGGG ACTATATCAT 960

CGAAAAAATG ACTGCTCTTG GTTTTGAGAT TATCAAACCA GACGGTGCCT TCTATATTTT 1020

TGCTAAAATT CCAGCGGGCT ACAATCAAGA CTCCTTTGCT TTTCTGAAGG ATTTTGCTCA 1080

GAAGAAGGCC GTTGCCTTTA TCCCTGGTGC AGCCTTTGGA CGTTACGGGG AAGGCTACGT 1140

CCGCCTATCT TATGCAGCCA GCATGGAGAC TATCAAAGAA GCCATGAAAC GACTTGAGGA 1200

GTACATGAGA GAAGCATGAT TCAGTCTATC ACGAGTCAAG GCTTGGTGCT TTACAATCGC 1260

AATTTTCGTG AGGATGACAA GCTCGTCAAA ATTTTTACAG AGCAGGTTGG CAAACGCATG 1320

TTTTTTGTCA AACACGCTGG TCAGTCTAAG CTGGCGCCTG TTATTCAGCC CTTGGTGCTG 1380

GCACGATTTC TCTTGCGAAT CAATGATGAC GGACTCAGTT ACATCGAAGA CTATCATGAG 1440

GTCATGACTT TTCCCAAGAT TAATAGTGAC CTCTTTGTCA TGGCCTATGC GACCTATGTG 1500

GCAGCTCTTG CAGATGCTAG TTTGCAGGAC AATCAGCAGG ATGCTCCCTT GTTTGCTTTT 1560

TTGCAAAAGA CTTTGGAGTT GATGGAAGCA GGCTTGGATT ATCAGGTTTT GACCAATATT 1620

TTTGAAATTC AAATTTTGAC TCGATTTGGA ATCAGCCTCA ATTTTAATGA GTGTGTCTTC 1680

TGCCATCGGG TTGGTCAGGC TTTTGACTTT TCTTTCAAAT ATGGAGCCTG CCTCTGTCCA 1740

GAGCATTATC ATGAGGATAA GAGACGTTGT CATCTCAATC CCAATATCCC CTATCTGCTC 1800

AATCAATTTC AAGCTATTGA TTTTGAGACT TTGGAGACCA TTTCGCTCAA GCCTGGAATC 1860

AAGCAAGAGC TACGCCAATT TATGGATCAA TTATATGAAG AGTACGTTGG GATTCACCTA 1920

AAATCAAAGA AATTTATTGA TTCCCTAGCA GACTGGGGAC AATTACTAAA AGAGGAAAAG 1980

AAATGAAAAA AATCGCAGTA GATGCCATGG GGGGCGATTA CGCACCTCAG GCCATTGTTG 2040

AGGGTGTCAA TCAAGCCCTA TCTGACTTTT CAGATATCGA GGTTCAACTT TACGGAGATG 2100

AAGCTAAAAT CAAGCAATAT CTGACAGCGA CAGAGCGCGT CAGCATTATC CATACGGATG 2160

AGAAGATTGA TTCGGATGAT GAACCTACGA GAGCTATTCG GAATAAGAAA AATGCCAGTA 2220

TGGTATTGGC AGCCAAGGCT GTCAAAGATG GTGAAGCAGA CGCTGTCCTT TCGGCTGGGA 2280

ATACAGGTGC CTTGTTGGCA GCAGGATTCT TCATCGTGGG TCGTATCAAG AATATCGACC 2340

GTCCTGGACT CATGTCTACC TTGCCTACCG TTGATGGAAA AGGTTTTGAC ATGCTAGACC 2400

TTGGTGCCAA TGCAGAAAAT ACAGCCCAGC ACCTCCATCA ATATGCGGTT CTAGGTTCCT 2460

TCTATGCTAA AAATGTCCGT GGCATTGCGC AACCACGCGT TGGTTTGCTC AACAACGGAA 2520

CAGAGAGTAG CAAGGGCGAC CCGCTTCGTA AGGAAACTTA TGAATTACTG GCGGCTGATG 2580

AAAGTTTGAA CTTTATCGGA AACGTGGAAG CGCGTGATTT GATGAATGGC GTTGCAGATG 2640

TTGTTGTGGC AGATGGTTTC ACGGGAAACG CTGTGCTCAA ATCCATCGAA GGGACAGCTA 2700

TGGGAATCAT GGGCTTGCTC AAGACAGCTA TTACAGGTGG TGGTCTTCGA GCGAAACTAG 2760

GTGCCCTCCT TCTCAAGGAC AGCCTCAGTG GTTTGAAAAA ACAGCTCAAT TATTCAGATG 2820

TTGGTGGAGC GGTCTTGTTT GGTGTTAAGG CACCTGTTGT CAAGACTCAT GGCTCAAGCG 2880

ATGCCAAGGC TGTTTATAGT ACGATTCGTC AGATCCGTAC CATGCTAGAA ACAGACGTGG 2940

TTGCCCAGAC TGCGCGTGAA TTTTCAGGAG AATAAAAGAG ATGACAGAAA AAGAAATTTT 3000

TGACCGTATT GTGACCATTA TCCAAGAGCG ACAGGGAGAG GACTTTGTCG TGACAGAATC 3060

CTTGAGTCTG AAAGACGATT TGGATGCGGA TTCTGTTGAC TTGATGGAGT TTATCTTGAC 3120

TCTGGAAGAT GAATTTAGTA TCGAAATCAG CGATGAAGAA ATTGACCAAC TCCAAAACGT 3180

AGGAGATGTG GTTAAAATCA TTCAAGGAAA ATAGCAATCG GAGTTCCAAG TCAACGGAAG 3240

TAGATGGTTT TTAGAAATGA GAAATATCGG ACAAGCTGGT AAAATCTTGG CTGACAGTGG 3300

TTATCAAGGG CTCATGAAGA TATATCCTCA AGCACAAACT CCACGTAAAT CCAGCAAACT 3360

CAAGCCGCTA ACAGTTGAAG ATAAAGCCTG TAATCATGCG CTATCTAAGG AGATAAGCAA 3420

GGTTGAGAAT ATCTTTGCCA AAGTAAAAAC GTTTAAAATG TTTTCAACAA CCTATCGAAA 3480

TCATCGTAAA CGCTTCGGAT TACGAATGAA TTTGATTGCT GGTATTATCA ATCATGAACT 3540

AGGATTCTAG TTTTGCAGGA AGTCTAATAG TAAAAAAGTG ATTAGAAAAC ATCTTTTTTA 3600

AAAATAGAGA TGATTTTGAA ACAAAAAAGC TAATTCAAGA CGTTTCGATG CCAATTCAAG 3660

ATTTGGATGA AAAAAATTAA TAGATACTGT TATACTAAAC TTGTCAAGTT TGTAACAAGA 3720

CAAATATTAA AAATAAAAAA GAGGTATTCG TTATGAATAC AAAAACGATG TCACAATTTG 3780

AAATTATGGA TACTGAGATG CTTGCTTGCG TTGAAGGTGG CGGATGCAAT TGGGGAGATT 3840

TTGCCAAAGC AGGTGTTGGA GGAGGAGCAG CACGAGGTCT TCAGCTAGGA ATTAAAACAA 3900

GAACATGGCA AGGTGCAGCA ACTGGTGCTG TGGGAGGAGC TATACTTGGA GGTGTGGCCT 3960

ATGCAGCGAC ATGTTGGTGG TAATTATGGA TTTTAAAAGT TTTATTATTG GTTTAGTAGT 4020

TGGTATATTT GGTCCTTATA TGGATGATTT AATTAGAAAA AAATTTTTAA AGTCTTCGGA 4080

GAAGAAAACA GAAAAATCTG TTAAAAAATA ATCAAAACTA TAAATGATGA ATCTGAATCA 4140

AAATTATTTT GCGCATGTAA AGAGGAGTCT TATAGTAACG AGTCAAAAAA GGAGTAACTA 4200

TGAATCGTAA TTTAGAACGG TGTTATCTAT TCTGACTAGG AATAGATCAT ACCAGAGGTA 4260

GCTTAGAAAT AGCAGAGACA TTAGAAATTG AAGTAATAAA TAGGATGTCG TAAGTGTTAC 4320

TATCAATGAT TTATTTGTTT CAAGCTTGCC TAGGGTGACA GTAAAAAATC AATTTCCTTT 4380

CAATAGCATA TTTTTAGTGG GCAGGACTCT TGTTCTGCCT ATTTTTTTAT CCAAAAAGTG 4440

CAGTTGGGAG GGAGATAGGC TCATTTGGGA AGGAAGTCCA GTTTTTGTTT AGTGATTGGG 4500

GTAAGATAGT TGTTATCAGA TGAGTTAATA CTCTTCGAAA ATCAAATTCA AACCACGTCA 4560

ACGTCGCCTT GCCGTATATA TGTGACTGAC TTCGTCAGTC CTATCTACAA CCTCAAAACA 4620

GTGTTTTGAG CAGCCTACGG CTAGTTTCCT AGTTTGCTCT TTGATTTTCA TTGAGTATTA 4680

GGGAAAAGGA GATGAATATG AAATTTGGGA AACGTCATTA TCGTCCGCAG GTGGATCAGA 4740

TGGACTGCGG TGTAGCTTCA TTAGCCATGG TTTTTGGCTA CTATGGTAGT TATTATTTTT 4800

TGGCTCACTT GCGAGAATTG GCTAAGACGA CCATGGATGG GACGACGGCT TTGGGCTTGG 4860

TCAAGGTGGC AGAGGAGATT GGTTTTGAGA CGCGAGCCAT TAAGGCAGAT ATGACGCTTT 4920

TTGACTTGCC GGATTTAACT TTTCCTTTTG TTGCCCATGT GCTTAAGGAA GGGAAATTGC 4980

TCCACTACTA TGTGGTGACT GGGCAGGATA AGGATAGCAT TCATATTGCC GATCCAGATC 5040

CCGGGGTGAA GTTGACTAAA CTGCCACGTG AGCGTTTTGA GGAAGAATGG ACAGGAGTGA 5100

CTCTTTTTAT GGCACCTAGT CCAGACTATA AGCCTCATAA GGAACAAAAA AATGGTCTGC 5160

TCTCTTTTAT CCCTATATTA GTGAAGCAGC GTGGCTTGAT TGCCAATATC GTTTTGGCAA 5220

CACTCTTGGT AACCGTGATT AACATTGTGG GTTCTTATTA TCTGCAGTCT ATCATTGATA 5280

CCTATGTGCC AGATCAGATG CGTTCGACAC TAGGGATTAT TTCTATTGGG CTAGTCATCG 5340

TCTACATCTT CCAGCAAATC TTGTCTTACG CTCAGGAGTA TCTCTTGCTT GTTTTGGGGC 5400

AACGCTTGTC GATTGACGTG ATTTTGTCCT ATATCAAGCA TGTTTTTCAC CTCCCTATGT 5460

CCTTCTTTGC GACACGCAGG ACAGGGGAGA TCGTGTCTCG TTTTACAGAT GCTAACAGTA 5520

TCATCGATGC GCTGGCTTCG ACCATCCTTT CGATTTTCCT AGATGTGTCA ACGGTTGTCA 5580

TTATTTCCCT TGTTCTATTT TCACAAAATA CCAATCTCTT TTTCATGACT TTATTGGCGC 5640

TTCCTATCTA CACAGTGATT ATCTTTGCCT TTATGAAGCC GTTTGAAAAG ATGAATCGGG 5700

ATACCATGGA AGCCAATGCG GTTCTGTCTT CTTCTATCAT TGAGGACATC AACGGTATTG 5760

AGACTATCAA GTCCTTGACC AGTGAAAGTC AGCGTTACCA AAAAATTGAC AAGGAATTTG 5820

TGGATTATCT GAAGAAATCC TTTACCTATA GTCGAGCAGA GAGTCAGCAA AAGGCTCTGA 5880

AAAAGGTTGC CCATCTCTTG CTTAATGTCG GCATTCTCTG GATGGGGGCT GTTCTGGTCA 5940

TGGATGGCAA GATGAGTTTG GGGCAGTTGA TTACCTATAA TACCTTGCTG GTTTACTTTA 6000

CTAATCCTTT GGAAAATATC ATCAATCTGC AAACCAAGCT TCAGACAGCG CAGGTTGCCA 6060

ATAACCGTCT AAATGAAGTG TATCTAGTAG CTTCTGAGTT TGAGGAGAAG AAAACAGTTG 6120

AGGATTTGAG CTTGATGAAG GGAGATATGA CCTTCAAGCA GGTTCATTAC AAGTATGGCT 6180

ATGGTCGAGA TGTCTTATCG GATATCAATT TAACCGTTCC CCAAGGGTCT AAGGTGGCTT 6240

TTGTGGGGAT TTCAGGGTCA GGTAAGACGA CTTTGGCCAA GATGATGGTT AATTTTTACG 6300

ACCCAAGTCA AGGGGAGATT AGTCTGGGTA GTGTCAATCT CAATCAGATT GATAAAAAAG 6360

CCCTGCGCCA GTACATCAAC TATCTGTCTC AACAGCCCTA TGTCTTTAAC GGAACGATTT 6420

TGGAGAATCT TCTTTTGGGA GCCAAGGAGG GGACGACACA GGAAGATATC TTACGGGCGG 6480

TCGAATTGGC AGAGATTCGA GAGGATATCG AGCGCATGCC ACTGAATTAC CAGACAGAAT 6540

TGACTTCGGA TGGGGCAGGG ATTTCAGGTG GTCAACGTCA GAGAATCGCT TTGGCGCGTG 6600

CTCTCTTGAC AGATGCGCCG GTCTTGATTT TGGATGAGGC GACTAGCAGT TTGGATATTT 6660

TGACAGAGAA GCGGATTGTC GATAATCTCA TTGCTTTGGA CAAGACCTTG ATTTTCATTG 6720

CTCACCGCTT GACTATTGCT GAGCGGACAG AGAAGGTAGT TGTCTTGGAT CAGGGCAAGA 6780

TTGTCGAAGA AGGAAAGCAT GCTGATTTGC TTGCACAGGG TGGCTTTTAC GCCCATTTGG 6840

TCAATAGCTA GAAAGAGGAG AGGATGAAAC CAGAATTTTT AGAAAGTGCG GAGTTTTATA 6900

ATCGTCGTTA CCATAATTTT TCCAGTAGTG TGATTGTACC CATGGCCCTT CTGCTTGTGT 6960

TTTTACTTGG CTTTGCAACT GTTGCAGAGA AGGAGATGAG TTTGTCCACT AGAGCTACTG 7020

TCGAACCTAG TCGTATCCTT GCAAATATCC AGTCAACTAG CAACAATCGT ATTCTTGTCA 7080

ATCATTTGGA AGAAAATAAG CTGGTTAAGA AGGGGGATCT TTTGGTTCAA TACCAAGAAG 7140

GGGCAGAGGG TGTCCAAGCG GAGTCCTATG CCAGTCAGTT GGACATGCTA AAGGATCAAA 7200

AAAAGCAATT GGAGTATCTG CAAAAGAGCC TGCAAGAAGG GGAGAACCAC TTTCCAGAGG 7260

AGGATAAGTT TGGCTACCAA GCCACCTTTC GCGACTACAT CAGTCAAGCA GGCAGTCTTA 7320

GGGCTAGTAC ATCGCAACAA AATGAGACCA TCGCGTCCCA GAATGCAGCA GCTAGCCAAA 7380

CCCAAGCCGA AATCGGCAAC CTCATCAGTC AAACAGAGGC TAAAATTCGC GATTACCAGA 7440

CAGCTAAGTC AGCTATTGAA ACAGGTGCTT CCTTGGCCGG TCAGAATCTA GCCTACTCTC 7500

TTTACCAGTC CTACAAGTCT CAGGGCGAGG AAAATCCCCA AACTAAGGTT CAGGCAGTTG 7560

CACAGGTTGA AGCACAGATT TCTCAGTTAG AATCTAGTCT TGCTACTTAC CGTGTCCAGT 7620

ATGCAGGTTC AGGTACCCAG CAAGCCTATG CGTCAGGGTT AAGCAGTCAA TTGGAATCCC 7680

TTAAATCCCA ACACTTGGCA AAGGTTGGTC AGGAATTGAC CCTTCTAGCC CAGAAAATTT 7740

TGGAGGCAGA GTCAGGTAAG AAGGTACAGG GAAATCTTTT AGACAAGGGG AAAGTTACGG 7800

CGAGTGAGGA TGGGGTGCTT CATCTTAATC CTGAGACCAG TGATTCTAGC ATGGTTGCAG 7860

AAGGTGCCCT ACTAGCCCAA CTTTATCCAT CTTTGGAAAG AGAAGGGAAA GCCAAACTCA 7920

CAGCTTATCT AAGTTCAAAA TATGTAGCAA GAATCAAGGT CGGTGATTCT GTTCGCTATA 7980

CTACGACTCA TGATGCCGGG AATCAACTTT TCCTAGATTC TACTATTACA AGTATTGATG 8040

CGACAGCTAC TAAGACTGAG AAAGGGAATT TCTTTAAAAT CGAGGCGGAG ACTAATCTAA 8100

CTTCGGAGCA GGCTGAAAAA CTTAGGTACG GGGTGGAAGG CCGCTTGCAG ATGATTACGG 8160

GCAAGAAAAG TTACCTACGT TATTATTTGG ATCAATTTTT GAACAAAGAG TAATGTTCGT 8220

GTTTTTAGAG TTAAATAATT TTTAAACTGT GAGAAAGATT CTTCTTGCAG TTTTTTCTTT 8280

ACAATTTTTG AAAAACATCT ACTATTTATT CGGTTAAATT CTTGTGTTTT TTGGTTTTTT 8340

GTGGTAAAAT GTGCTCAAGT AATACGAAAG GCGAACTTTA AAATGTCAAA ACAATTGATC 8400

TATTCGGGAA AAGCTAAAGA TATCTATACA ACTGAGGATG AAAATCTTAT TATTTCAACT 8460

TACAAGGACC AGGCGACTGC TTTCAACGGT GTCAAGAAGG AGCAGATTGC AGGTAAGGGA 8520

GTCTTGAATA ATCAGATCTC ATCTTTTATT TTTGAGAAAT TAAATGTGGC TGGTGTGGCG 8580

ACTCACTTTG TGGAGAAACT TTCAGACACG GAACAACTCA ATAAAAAGGT TAAGATTATT 8640

CCTTTGGAAG TCGTGCTCCG CAACTATACT GCTGGTTCCT TTTCAAAACG TTTTGGTGTG 8700

GATGAGGGAA TCGCCTTGGA GACTCCGATT GTCGAATTTT ACTACAAAAA TGATGATTTG 8760

GATGATCCAT TTATCAATGA TGAGCATGTG AAATTCCTAC AGATTGCGGG TGACCAGCAG 8820

ATTGCCTACT TGAAGGAAGA AACGCGTCGT ATCAATGAAC TATTGAAAGT CTGGTTTGCT 8880

GAGATTGGGC TTAAATTGAT TGACTTTAAG CTAGAGTTCG GTTTTGACAA GGATGGCAAG 8940

ATTATCTTGG CAGACGAATT TTCACCAGAT AACTGCCGCT TGTGGGACGC TGATGGCAAC 9000

CACATGGATA AGGATGTTTT CCGTAGAGGA TTGGGAGAAC TAACCGACGT TTATGAGATT 9060

GTTTGGGAAA AGTTGCAGGA ATTGAAATAA TCTGTTTGCA ACGGAAAACC TTCGTCTCTC 9120

AACTAAAAGG ACTCAGGCTG AAAAGGTCCC CCAGACCTTT TCACTCTGTA GAGAACTAGG 9180

TGAACTAACA GATGTTTACG AAATTGTCTG GGAAAAGTTG CAGGGTTTAA AATAACAACC 9240

TCAAGGCTGT TTGGGAATAT TGCAAGAGCT GAAATAAAGG AATAAGAATT GATGGATAAA 9300

CGTATTTTTG TTGAAAAAAA GGCTGATTTT CAGGTCAAGT CAGAGAGTTT GGTTAGAGAG 9360

CTCCAGCACA ACTTGGGACT GTCAAGCTTG AAAAGTATTC GTATTGTGCA AGTATATGAT 9420

GTATTTGACT TGGCTGAGGA CTTGTTTGCA CCTGCAGAGA AGCACATTTT CTCTGAGCAG 9480

GTAACCGACC ATGTTTTAGA TGAAGTATCT GTGCAGGCGG ATCTTGCTAA CTATGCTTTC 9540

TTTGCCATTG AAAGTCTGCC AGGGCAGTTT GACCAGCGTG CAGCTTCGTC ACAGGAAGCC 9600

TTGCTTTTGT TGGGAAGTTC GAGTGACGTG ACAGTCAACA CAGCCCAACT TTACTTGGTG 9660

AATAAAGATA TTGATGCGAC TGAGTTGGAA GCTGTCAAAA ACTACCTGCT CAATCCAGTT 9720

GATTCTCGTT TCAAGGATAT CACGACAGGG ATTGCCAAGC AGGAGTTTTC AGAGTCAGAC 9780

AAGACCATTC CCAAATTGAC TTTCTTTGAA AGCTATGCAG CAGAAGACTT TGCTCGCTAC 9840

AAGGCCGAAC AAGGGATGGC CATGGAAGTG GATGATTTGC TCTTTATCCA AGACTACTTT 9900

AAGTCAATCG GGCGCGTGCC AACTGAGACT GAACTCAAGG TTTTGGACAC TTACTGGTCT 9960

GACCACTGCC GTCATACGAC TTTTGAGACA GAGTTGAAAC ACATCGACTT TTCAGCTTCT 10020

AAATTTCAAA AGCAATTGCA GTCAACCTAT GACAAGTATA TTGCCATGCG CGAGGAATTA 10080

GGTCGGTCTG AAAAACCACA AACCTTGATG GATATGGCGA CTATTTTCGG TCGTTATGAG 10140

CGTGCTAATG GACGATTGGA TGATATGGAA GTCTCTGACG AAATCAATGC CTGCTCAGTT 10200

GAAATTGAAG TGGACGTTGA TGGTGTCAAG GAACCTTGGC TCCTCATGTT TAAAAACGAA 10260

ACCCACAACC ATCCAACAGA AATTGAGCCA TTTGGTGGAG CGGCTACCTG TATTGGTGGA 10320

GCTATTCGTG ATCCGTTGTC AGGCCGTTCC TATGTTTACC AAGCCATGCG TATTTCAGGT 10380

GCTGGTGATA TTACAGCACC GATTTCGGAA ACTCGCGCTG GGAAATTGCC ACAACAAGTC 10440

ATTTCTAAAA CAGCAGCTCA TGGTTATTCT TCATATGGTA ACCAGATTGG GCTTGCAACA 10500

ACCTACGTTC GTGAATACTT CCACCCAGGC TTTGTAGCTA AACGTATGGA ACTTGGTGCC 10560

GTTGTTGGTG CGACTCCCAA GGGCAATGTT GTCCGTGAAA AACCTGAAGC AGGTGATGTG 10620

ATCATCCTTC TCGGAGGCAA AACAGGTCGT GATGGTGTCG GTGGTGCGAC GGGCTCTTCT 10680

AAGGTTCAAA CAGTTGAGTC TGTAGAGACT GCTGGTGCTG AGGTTCAAAA AGGAAATGCC 10740

ATCGAAGAAC GCAAGATTCA GCGCCTCTTC CGTAATGGCA ATGTCACTCG TCTGATCAAG 10800

AAGTCCAATG ACTTTGGGGC AGGCGGTGTC TGTGTGGCTA TCGGTGAATT GGCAGACGGT 10860

CTTGAAATCG ACCTCAACAA GGTGCCTCTT AAATACCAGG GCTTGAATGG TACAGAAATT 10920

GCCATCTCTG AATCACAAGA ACGGATGGCG GTCGTGGTTC GTCCTGAAGA TGTGGATGCC 10980

TTCGTTGCCG AATGTAACAA AGAAAATATT GATGCTGTTG TGGTGGCGAC AGTAACTGAA 11040

AAACCAAATC TTGTCATGCA CTGGAATGGT GAGACAATCG TTGACTTGGA GCGTCGTTTC 11100

CTTGACACCA ATGGTGTGCG CGTGGTTGTC GATGCCAAAG TTGTGGACAA GGATGTCAAA 11160

CTCCCAGAAG AGCGTCAAAC ATCTGCTGAA ACACTGGAAT CAGATACCCT TACGGTTCTA 11220

TCTGACCTCA ACCATGCAAG TCAAAAAGGA TTACAGACTA TCTTTGACTG CTCTGTTGGA 11280

CGCTCAACGG TTAATCACCC ACTTGGTGGT CGTTACCAAC TCACACCAAC TGAGGCATCT 11340

GTGCAGAAAT TGCCAGTTCA ACACGGTGTG ACTCATACTG CGTCGGTCAT TGCTCAAGGT 11400

TTCAACCCAT ATGTAGCTGA ATGGTCTCCA TACCACGGTG CTGCTTATGC GGTTATCGAA 11460

GCAACTGCTC GTTTGGTGGC TGCTGGTGCC AACTGGTTCA AGGCTCGTTT CTCTTACCAA 11520

GAGTATTTCG AGCGTATGGA TAAACAAGCA GAGCGTTTCG GTCAGCCAGT AGCTGCTCTT 11580

CTAGGTTCTA TTGAAGCACA AATTCAGCTT GGCTTGCCAT CTATCGGTGG TAAGGACTCC 11640

ATGTCTGGTA CCTTTGAAGA ATTGACCGTT CCGCCAACCT TGGCTGCCTT TGGGGTGACG 11700

ACGGCAGATA GCCGTAAGGT GCTCTCTCCA GAATTTAAAG CTGTTGGGGA AAATATCTAC 11760

TACATCCCAG GTCAAGCCCT CTCTGCAGAG ATTGATTTTG ACTTGATTAA GAAAAATTTT 11820

GCTCAGTTTG AAGCCATCCA AGCTGACCAT AAAGTGACAT CTGCATCAGC TGTCAAATAC 11880

GGTGGTGTAG TTGAAAGTTT GGCTCTTGCT ACCTTTGGAA ACTATATTGG TGCAGAGGTG 11940

ACCTTGCCTG AACTTGAAAC AGCTTTGACA GCTCAATTAG GCGGCTTTGT CTTCACATCT 12000

CCTGAAGAAA TTGCTGGAGT AGAGAAGGTT GGACAAACGA AAGCAGACTT TACACTGACT 12060

GTCAACGGTG TGAAGCTAGA TGGACACAAG CTTGACAGTG CATTTCAAGG GACATTGGAA 12120

GAAGTTTACC CAACAGAATT TACCCAAGCG AAAGAACTAG AAGAAGTACC AGCTGTGGCA 12180

TCAGATGTTG TGATTAAAGC CAAAGAAAAG GTTGAAAAAC CTGTGGTTTA CATCCCAGTC 12240

TTTCCAGGAA CCAACTCAGA ATATGATTCA GCTAAGGCCT TCGAAAAAGA AGGTGCAGAG 12300

GTCAATTTGG TGCCATTCGT GACCTTGAAT GAAGAAGCTA TTGTCAAGTC AGTTGAAACT 12360

ATGGTTGACA ATATCGACAA GACTAATATT CTCTTCTTTG CTGGTGGATT CTCGGCTGCG 12420

GATGAACCAG ATGGTTCAGC TAAGTTTATC GTCAATATCC TGCTTAATGA AAAAGTGCGT 12480

GTGGCTATTG ATAGCTTTAT CGCCCGTGGT GGTTTGATTA TCGGTATTTG TAATGGATTC 12540

CAAGCCTTAG TCAAATCGGG TCTCCTACCC TACGGAAACT TTGAAGCTGC TAACAGTACT 12600

AGCCCAACCC TCTTCTACAA TGATGCCAAC CAACACGTGG CCAAGATGGT GGAAACTCGC 12660

ATTGCCAATA CCAACTCACC ATGGTTGGTT GGTGTGCAAG TGGGCGATAT CCACGCTATT 12720

CCTGTTTCGC ACGGTGAAGG GAAGTTTGTC GTGACGGCTG AGGAATTTGC AGAGCTCCGT 12780

GACAATGGAC AAATTTTCAG CCAATACGTT GACTTTAACG GTAAACCAAG TATGGATTCT 12840

AAGTACAATC CGAATGGTTC TGTCCATGCC ATCGAAGGAA TTACCAGCAA GAATGGTCAA 12900

ATCATCGGTA AGATGGGCCA CTCAGAACGT TATGAGGATG GTCTTTTCCA AAATATCCCA 12960

GGCAATAAAG ACCAACACCT GTTCGCATCA GCGGTTAAAC ATTTCACTGG AAAATAAGAC 13020

TTACAGATTT TCTAATAGAT AGTATCAGTA ATGTAAAAGT CATGTAAATC TAGCTCTTGA 13080

TGATTACAAA TGAAAATTAG GTATAAAAAA TGACATACGA AGTAAAATCT CTTAATGAAG 13140

AATGTGGTGT TTTCGGTATT TGGGGACATC CAGATGCTGC TAAGTTGACC TATTTTGGAC 13200

TCCACAGTCT TCAACACCGT GGTCAGGAGG GGGCAGGAAT CCTCTCCAAT GATCAAGGAC 13260

AACTGAAGCG CCATCGTGAC ATGGGGCTTT TATCAGAAGT TTTCAGAAAT CCAGCTAATT 13320

TGGATAAATT GACAGGAGCT GGTGCGATTG GGCATGTGCG TTATGCGACT GCTGGCGAAG 13380

CTTCTGTAGA TAACATCCAG CCCTTCCTCT TCCGTTTTCA CGATATGCAG TTTGGTTTGG 13440

CTCATAATGG AAATCTGACC AATGCAGCCT CTCTCAAGAA AGAACTGGAA CAAAGAGGAG 13500

CAATTTTCAG CGCGACTTCG GACTCTGAAA TCTTGGCTCA CCTCATTCGT CGCAGTCATA 13560

ATCCTAGCCT GATGGGCAAA ATCAAGGAAG CGCTCAGCCT TGTCAAAGGT GGTTTTGCCT 13620

ATATCTTGCT GTTTGAGGAC AAGTTGATTG CGGCTCTTGA CCCAAATGGA TTCCGACCGC 13680

TTTCGATTGG TAAAATGGCT AATGGAGCAG TTGTTGTATC TTCTGAAACC TGTGCTTTTG 13740

AGGTCATTGG TGCCGAGTGG ATTCGTGATT TGAAGCCAGG TGAGATTGTG ATCATTGATG 13800

ACGAGGGCAT TCAGTATGAC AGCTATACAG ATGATACCCA GTTGGCGGTT TGTTCTATGG 13860

AGTATATCTA CTTTGCTCGC CCTGATTCTA ATATCCACGG TGTCAATGTC CATACGGCAC 13920

GTAAGAGAAT GGGAGCGCAA TTGGCGCGAG AATTTAAGCA TGAGGCAGAT ATTGTAGTTG 13980

GTGTGCCCAA TTCTTCCCTA AGCGCGGCTA TGGGATTTGC GGAAGAATCA GGCTTACCAA 14040

ATGAAATGGG TCTGATCAAA AACCAATACA CCCAGCGAAC TTTTATCCAA CCGACTCAAG 14100

AATTGCGGGA GCAAGGAGTG CGGATGAAAC TGTCTGCTGT TTCGGGTGTT GTCAAAGGCA 14160

AACGTGTGGT CATGGTGGAT GATTCCATTG TACGTGGAAC AACCTCTCGT CGTATCGTTC 14220

AGCTCTTGAA AGAAGCGGGT GCGACTGAGG TTCACGTTGC CATTGGAAGT CCTGCACTAG 14280

CGTATCCATG TTTCTACGGG ATTGATATCC AGACCCGTCA GGAGCTGATT GCAGCCAATC 14340

ATACGGTCGA AGAAACTCGC CAAATCATTG GTGCGGACAG TCTGACTTAT CTTTCAATTG 14400

ATGGCTTGAT TGAGTCGATT GGTATCGAAA CAGATGCGCC GAACGGTGGT CTCTGTGTCG 14460

CTTACTTTGA CGGTGACTAC CCAACGCCTC TTTATGACTA CGAAGAAGAC TATCGTAGAA 14520

GTTTGGAAGA AAAGACCAGT TTTTACAAGT AGGCGACAGA TTCTCCATTA AAGAAAAGGA 14580

AAAAATAAAT GACAAATAAA AATGCATATG CCTCACGTCT CACTACTGAC TAAAGGCTTA 14640

AGCATTTAGT CAGTAGACGC TTTGTCCTAT AGGATCAAAG CTAGAGCCCT GACTAGTATT 14700

TTTAGATAAA AAGATGGTTT ATCTAAAAAT ACGTCGCAGT CTTTCTCAAA AAAAGAAAAG 14760

GAAAAATAAA ATGGCAAATA AAAATGCGTA CGCTCAATCT GGTGTGGATG TTGAAGCGGG 14820

TTATGAAGTT GTTGAACGGA TTAAAAAGCA CGTGGCCCGT ACGGAGCGTG CAGGTGTCAT 14880

GGGAGCTCTT GGTGGCTTTG GTGGTATGTT TGACCTTTCC AAGACTGGGG TTAAAGAACC 14940

CGTCTTGATT TCAGGGACTG ACGGTGTCGG AACCAAGCTC ATGTTGGCTA TCAAGTACGA 15000

CAAGCACGAT ACCATCGGGC AGGACTGTGT GGCCATGTGT GTCAACGACA TCATTGCTGC 15060

AGGTGCGGAA CCCCTCTATT TTCTCGACTA CGTAGCGACA GGGAAGAATG AACCAGCTAA 15120

GCTAGAACAA GTGGTTGCTG GTGTGGCAGA AGGTTGTGTG CAGGCTGGTG CTGCCCTCAT 15180

CGGTGGGGAA ACGGCTGAAA TGCCGGGCAT GTACGGCGAA GACGACTATG ACTTGGCTGG 15240

TTTTGCGGTC GGTGTGGCTG AAAAATCTCA AATCATTGAC GGTTCAAAGG TGGTAGAGGG 15300

AGATGTTCTT CTCGGACTTG CTTCAAGTGG GATTCACTCA AATGGTTACT CTTTGGTTCG 15360

TCGTGTCTTT GCGGATTACA CAGGTGAGGA AGTCCTACCA GAATTGGAAG GCAAGAAACT 15420

TAAGGAAGTT CTACTTGAGC CGACTCGTAT CTATGTCAAG GCTGTCTTGC CGCTCATCAA 15480

AGAAGAGTTG GTCAACGGCA TTGCCCACAT CACAGGTGGT GGCTTTATCG AAAATGTCCC 15540

TCGTATGTTT GCAGATGACC TAGCTGCTGA AATTGATGAA AGTAAAGTTC CAGTGCTTCC 15600

AATTTTCAAA ACCCTTGAAA AATACGGTCA GATTAAACAC GAAGAAATGT TTGAAATCTT 15660

CAATATGGGT GTGGGACTTA TGTTGGCGGT CAGCCCTGAA AATGTAGAGC GTGTAAAAGA 15720

ATTGTTGGAT GAAGCAGTCT ATGAAATTGG TCGCATCGTC AAGAAAGAAA ACGAAAGTGT 15780

CATTATCAAA TGAAAAAAAT AGCGGTTTTT GCCTCTGGTA ATGGCTCAAA TTTTCAGGTG 15840

ATTGCCGAAG AATTTCCAGT GGAGTTTGTC TTTTCAGACC ATCGTGATGC CTATGTGCTT 15900

GAGCGTGCAA AGCAGCTCGG CGTTCTGTCC TATGCTTTTG AACTCAAGGA GTTTGAGAGC 15960

AAGGCAGACT ACGAAGCAGC CCTTGTCGAA CTCTTGGAAG AACACCAGAT TGACTTGGTT 16020

TGCCTAGCAG GCTACATGAA AATCGTTGGA CCAACCTTAT TGTCGGCTTA TGAAGGTCGG 16080

ATTGTCAACA TTCATCCAGC CTACTTGCCA GAATTTCCAG GAGCTCATGG GATTGAGGAT 16140

GCTTGGAATG CTGGCGTGGG TCAGTCTGGT GTGACCATTC ACTGGGTGGA TTCGGGTGTG 16200

GATACAGGCC AGGTCATCAA ACAGGTTCGT GTGCCACGAC TAGCTGATGA TACCATTGAC 16260

AGATTTGAAG CTCGCATCCA TGAAGCAGAG TACAGGCTGT ATCCGGAAGT AGTGAAGGCT 16320

CTATTTACAG ATTGACTTTT TGATGATTCA TATGATATCT TTGATTTTAA ATTGGAGTCA 16380

GTGTTTGTTG AAGACGGCTT CAAACGGAGG TATTTGTAAT GTTAGAATCT AAAAAAACAA 16440

CTCGATATGT ATTTTATGTC TATCTGATGT TATTAACTTG GGGAATCTTA TTTAAGTTTG 16500

AAACAAATCC TGAATTTATA GCATTTTTCT TAGCTCCAAG GTATATCAAT TGGATTCCAT 16560

TTTCAGAACC ACTAATAGTC GATGGAAAAA TTGTTTTTGC TGAAATGTTA TTTAATCTGA 16620

TTTTCTTTAT TCCATTAGGT GTTTGTTTCC CTTTGATAAA AACTAATTTA TCTAGTTTAA 16680

GAATAGTCGG GACAGGTTTC TTGATTAGTT TATTGTTTGA GTGCTTACAG TATATTTTAG 16740

CAATAGGTAT AACAGATATA ACGGATTTGA CTTTAAATAC GCTAGGTGTC TGTGTAGGCT 16800

TACTGATTTA TCAAATTTTT ATAAGAGTGT TCAAATCACA GACTAGAAAA TGGATCAATA 16860

TCTTAGGTAT GCTTAGCCTT GGTTTTGCTT ATCTTGTTTT ACTGTTACTG CATTTACTTA 16920

GTGTTTAACT AATGATTAAA AAGGAGAATA TAATGACTAA ACGCGTCTTA ATCAGCGTCT 16980

CAGACAAAGC GGGCATTGTT GAATTTGCCC AAGAACTCAA AAAACTTGGT TGGGAGATTA 17040

TCTCAACAGG TGGAACTAAG GTTGCCCTTG ATAATGCTGG GGTGGATACC ATTGCTATCG 17100

ATGATGTGAC TGGTTTCCCA GAAATGATGG ACGGTCGTGT GAAGACCCTC CACCCAAATA 17160

TCCACGGAGG GCTTCTCGCT CGTCGTGACT TGGATAGCCA CTTGGAAGCG GCTAAGGACA 17220

ACAAGATTGA GCTCATTGAC CTTGTGGTGG TCAACCTTTA CCCATTTAAG GAAACTATCC 17280

TTAAACCAGA TGTGACTTAT GCTGATGCAG TTGAAAATAT CGATATTGGT GGGCCATCTA 17340

TGCTTCGTTC AGCAGCGAAA AATCATGCCA GTGTTACAGT TGTGGTAGAT CCTGCTGACT 17400

ACGCTGTGGT TTTGGATGAA TTGGCAGCAA ACGGCGAAAC CTCTTATGAA ACTCGCCAAC 17460

GTTTAGCAGC CAAAGTATTT CGTCACACAG CGGCTTATGA CGCCTTGATT GCAGAATACT 17520

TCACAGCTCA AGTGGGTGAA AGCAAGCCTG AAAAACTCAC TTTGACTTAT GACCTCAAGC 17580

AACCAATGCG TTACGGTGAG AATCCTCAAC AAGACGCGGA CTTTTACCAG AAAGCTTTGC 17640

CTACAGACTA CTCCATTGCT TCAGCCAAAC AGCTCAACGG GAAAGAATTG TCATTTAATA 17700

ATATCCGTGA TGCAGATGCT GCTATCCGTA TCATCCGTGA CTTCAAAGAT AGTCCAACCG 17760

TTGTGGCTCT CAAACACATG AATCCATGTG GAATTGGTCA AGCTGATGAC ATCGAGACTG 17820

CTTGGGACTA CGCTTATGAG TCTGACCCAG TGTCTATCTT TGGTGGGATT GTCGTCCTCA 17880

ACCGTGAGGT GGATGCTGCG ACAGCTGAGA AGATGCACGG CGTTTTCCTC GAAATCATCA 17940

TTGCACCAAG CTATACGGAT GAAGCGCTAG CCATTTTGAT CAATAAAAAG AAAAACTTGC 18000

GTATCCTTGC CTTGCCATTT AATGCTCAAG AGGCTAGCGA AGTGGAAGCA GAATACACAG 18060

GTGTAGTCGG TGGACTTCTC GTGCAAAATC AAGACGTGGT CAAGGAAAGC CCAGCTGACT 18120

GGCAAGTGGT GACTAAACGT CAGCCAACTG AGACAGAAGC GACTGCTCTT GAGTTCGCTT 18180

GGAAGGCTAT CAAGTACGTC AAATCAAATG GTATTATCGT GACCAACGAC CACATGACAC 18240

TTGGTGTTGG TCCAGGTCAA ACCAACCGTG TGGCTTCTGT TCGCCTTGCC ATTGACCAAG 18300

CCAAAGATCG TCTGGACGGG GCGGTCCTTG CTTCAGATGC CTTCTTCCCA TTTGCGGATA 18360

ACGTGGAAGA AATCGCCAAA GCAGGAATTA AGGCCATCAT CCAGCCCGGT GGCTCTGTCC 18420

GTGACCAAGA ATCCATCGAA GCAGCGGATA AATACGGCTT GACTATGGTC TTTACAGGTG 18480

TGAGACATTT TAGACATTAA GAAGATAAAA GGGAAGAAAA CAGTTTCTTT CCTTTTTTGG 18540

CTTAAAATAC TAACTGAAAC AAGATTAAAA CGAACTTTTT TGATATAATG TTGGTAAATA 18600

ATTCGCAAAA GAGGTTGAGG AATGAAACTG CTTGTTGTCG GTTCTGGTGG TCGTGAGCAT 18660

GCGATTGCTA AAAAGTTACT TGAATCAAAA GACGTGGAAA AAGTCTTTGT AGCTCCTGGG 18720

AATGATGGGA TGACTCTGGA TGGTTTGGAA TTGGTAAATA TCTCTATTTC CGAACATTAT 18780

AAATTGATTG ACTTCGCAAA GACCAATGAT GTTGCTTGGA CCTTTATCGG TCCAGATGAT 18840

GCCCTTGCTG CTGGTATCGT GGATGATTTT AACCAAGCTG GACTTAAGGC CTTTGGTCCG 18900

ACTAGGGCTG CAGCGGAGCT GGAGTGGTCC AAGGATTTCG CCAAGGAAAT CATGGTCAAA 18960

TACGGCGTTC CGACAGCAAC ATATGGCACA TTTTCAGATT TCGAGGAAGC CAAAGCCTAT 19020

ATCGAAAAGC ATGGTGCGCC TATCGTAGTC AAGGCGGATG GCTTGGCACT TGGGAAGGGT 19080

GTCGTCGTTG CTGAGACGGT TGAGCAAGCG GTCGAAGCCG CTCATGAGAT GCTTTTGGAC 19140

AATAAATTTG GTGACTCAGG TGCGCGCGTG GTTATTGAGG AATTCCTTGA AGGAGAGGAA 19200

TTTTCACTCT TTGCCTTTGT CAATGGTGAT AAGTTCTACA TCATGCCAAC GGCTCAGGAC 19260

CACAAACGTG CCTATGATGG CGACAAAGGG CCTAACACGG GTGGTATGGG TGCCTATGCG 19320

CCAGTCCCAC ACTTACCACA GAGTGTAGTT GATACAGCGG TTGACACCAT TGTCAAGCCA 19380

GTTCTAGAAG GGGTGATTAA AGAAGGTCGC CCTTATCTGG GAGTTCTTTA CGCAGGGCTT 19440

ATCCTGACAG CTGATGGACC GAAAGTCATT GAGTTCAACG CTCGGTTCGG AGATCCAGAA 19500

ACTCAGATTA TCTTGCCTCG CTTGACCTCT GACTTTGCTC AAAATATCAC AGATATCCTG 19560

GATAGCAAGG AGCCAAATAT CATGTGGACG GACAAGGGTG TGACTCTGGG TGTGGTTGTC 19620

GCATCCAAGG GCTACCCGCT AGACTATGAA AGGGGCGTTG AGTTGCCAGC CAAGACAGAA 19680

GGCGATGTCA TCACCTACTA TGCAGGGGCT AAGTTTGCGG AAAATAGCAG AGCACTGCTC 19740

TCAAACGGCG GACGAGTTTA TATGCTCGTT ACCACAGCAG ATACCGTCAA AGAAGCCCAA 19800

GCCAGCATAT ACCAAGAACT ATACCAACAA AAAATAGAAG GACTCTTCTA CCGAACAGAT 19860

ATCGGAAGCA AGGCAATTAA GTAAAGATAT AAGAATAACG CGCCGTAGTC GCCAAACACG 19920

ATAATGGTCG TCGTGGTGAA AAGACCAGAA CAGTGAATGT TCTGGTCAGG GGGAAACTTG 19980

GAGACCTTAG GCTCAAAGTT TAGGAATGAA ACCGAAGGTT TGCTTCCGCC TCCATCACCT 20040

AAGACCATTA TCAAAAAGAA AAATAAAAAT TCACAAAATA CGTTAATGAT CGTATGGTTT 20100

GCGAGCGTTA GCGAGCTAAT ATAGAACAAT CACCGCCGTT GTGAAAGAAC GATTGGATGA 20160

TAATCCAATC GTTCAGGGAA ATTGGAAGAC CTTGGGTTTC CAATTTAGGC ATGAGACACC 20220

TTTGGTGGCT GCTGCCGTCC CTCACAAGCT AAGGTGATTG TTGAAAAAGA GGAAAAAGGA 20280

GAAGAAATGA AACCAGTAAT TTCCATCATC ATGGGCTCAA AATCCGACTG GGCAACCATG 20340

CAAAAAACAG CAGAAGTCCT AGACCGCTTC GGTGTAGCCT ACGAAAAGAA AGTTGTTTCC 20400

GCACACCGTA CACCAGACCT CATGTTCAAA CATGCAGAAG AAGCCCGTAG TCGTGGCATC 20460

AAGATCATCA TCGCAGGTGC TGGTGGCGCA GCGCATTTGC CAGGCATGGT AGCTGCCAAA 20520

ACAACCCTTC CAGTCATTGG TGTGCCAGTC AAGTCTCGTG CTCTTAGTGG AGTGGATTCA 20580

CTCTATTCTA TCGTTCAGAT GCCGGGTGGG GTGCCTGTTG CGACCATGGC TATCGGTGAA 20640

GCTGGAGCGA CTAACGCAGC TCTCTTTGCC CTCCGTCTCC TCTCTGTAGA AGATAAGTCC 20700

ATTGCGGATG CACTTGCCAA CTTTGCTGAA GAACAAGGAA AAATCGCAGA GGAGTCGTCA 20760

AATGAGCTCA TCTAAAACAA TCGGAATTAT CGGTGGCGGT CAACTGGGTC AGATGATGGC 20820

CATTTCTGCT ATCTACATGG GCCACAAGGT TATCGCGCTG GATCCTGCGG CGGATTGCCC 20880

GGTCTCTCGT GTGGCGGAAA TCATTGTGGC ACCTTATAAC GATGTAGACG CCCTCCGTCA 20940

GTTGGCAGAC CGTTGCGATG TCCTCACTTA TGAGTTTGAA AATGTCGACG CTGACGGTTT 21000

GGATGCCGTT ATCAAGGATG GACAACTCCC TCAAGGAACA GATCTGCTCC GCATTTCGCA 21060

AAATCGTATT TTTGAAAAGG ACTTTTTGTC AAACAAGGCT CAAGTCACTG TGGCACCCTA 21120

CAAGGTCGTG ACTTCTAGCC TAGACTTGGC AGATATCGAC TTGTCGAAAA ACTATGTCCT 21180

CAAGACTGCG ACTGGTGGCT ACGATGGTCA TGGACAAAAG GTTATTCGTT CAGAAGCAGA 21240

CTTGGAAGCA GCCTATGCGC TAGCAGACTC AGCAGACTGC GTCTTGGAAG AATTTGTCAA 21300

CTTTGACCTT GAGATTTCTG TCATCGTGTC AGGAAATGGC AAGGAGGTGA CGTTTTTCCC 21360

AGTTCAGGAA AATATCCACC GCAACAATAT CCTGTCTAAG ACCATCGTAC CAGCCCGCAT 21420

TTCTGAAAGT CTAGTAGACA AGGCTAAAGC TATGGCAGTG CGAATCGCAG AACAACTCAA 21480

CTTGTCTGGA ACTCTCTGTG TGGAAATGTT TGCGACAGCT GATGACATCA TTGTCAATGA 21540

AATCGCCCCA CGACCACATA ACTCTGGGCA CTATTCTATT GAAGCCTGTG ATTTCTCTCA 21600

GTTTGACACC CATATTCTGG GTGTTCTCGG AGCACCATTA CCAGTCATCA AACTCCATGC 21660

GCCAGCCGTT ATGCTTAATG TCCTCGGTCA GCATGTCGAG GCTGCTGAAA AATATGTCAC 21720

AGAAAATCCA AGCGCCCACC TCCACATGTA TGGTAAAATA GAAGCAAAGC ATAATCGTAA 21780

GATGGGACAT GTGACTTTGT TTAGTGATGT GCCGGATAGT GTGGAAGAGT TTGGGGAAGG 21840

GATTGATTTT TAGGACAAGT CTATGATACA AATTATCGTT AATACATTTA TTGAAAAGTA 21900

TAAGACTGGA GCAGTTGTTG AAGTGTTGTA TGCCAGTGCT GACCAAGATA AGGTACAAGC 21960

TAAATATGAA GAACTAGCTG CACAATACCC CGAAAATTAT TTAGCTATCT ATAATGTACC 22020

GCTGGATACG GATTTGAATA CACTAGATCA TTACCCGTCT GTGTTTATTG GAAAAGAGGA 22080

GTTTGAGTAG AAATCTTGGT TTACCTAGAT AGCTTATTCC CAACAGCTTA AGAAGAAAGG 22140

AAAAATTAAC ACATGATCAA CCGTTACTCT CGCCCTGAGA TGGCGAATAT TTGGAGTGAA 22200

GAAAATAAAT ACCGTGCTTG GCTTGAGGTG GAAATCCTCT CTGACGAGGC ATGGGCTGAG 22260

TTGGGGGAAA TCCCTAAGGA AGATGTGGCT TTGATTCGCA AGAAGGCGGA CTTTGACATC 22320

GACCGTATTT TGGAAATTGA GCAGGAGACG CGCCACGATG TGGTGGCTTT CACGCGTGCG 22380

GTTTCTGAGA CTCTTGGTGA AGAGCGCAAG TGGGTTCACT ATGGGTTAAC TTCTACTGAC 22440

GTGGTGGATA CTGCTTATGG TTACCTCTAC AAGCAGGCCA ACGACATCAT CCGTCGTGAC 22500

CTTGAAAACT TCACTAATAT CATCGCTGAC AAGGCCAAGG AGCACAAGTT CACCATCATG 22560

ATGGGGCGTA CTCATGGTGT GCACGCTGAG CCGACAACCT TTGGTCTTAA ATTAGCAACT 22620

TGGTACAGCG AAATGAAACG CAATATCGAG CGCTTCGAGC ATGCGGCTGC TGGTGTAGAA 22680

GCTGGTAAGA TTTCTGGTGC GGTTGGGAAC TTTGCCAATA TCCCACCATT TGTAGAGGAG 22740

TATGTCTGCG ATAAACTTGG CATCCGTGCC CAAGAAATCT CTACACAAGT CCTTCCTCGT 22800

GACCTTCACG CTGAGTACTT TGCGGTTCTT GCCAGCATTG CGACTTCAAT CGAACGTATG 22860

GCGACTGAGA TTCGTGGTCT ACAAAAATCT GAGCAACGCG AAGTAGAAGA GTTCTTTGCT 22920

AAAGGGCAAA AAGGGTCTTC AGCAATGCCT CACAAACGCA ACCCAATCGG TTCTGAAAAT 22980

ATGACTGGTC TGGCGCGTGT CATTCGTGGT CACATGATTA CGGCTTATGA AAACGTCGCT 23040

CTCTGGCATG AACGCGATAT TTCTCACTCA TCAGCTGAGC GTATCATCAC ACCAGATACG 23100

ACCATTTTGA TTGACTACAT GCTCAACCGT TTTGGAAATA TCGTCAAGAA CTTGACAGTC 23160

TTCCCAGAAA ATATGATCCG AAACATGAAC TCGACTTTTG GTCTTATCTT TAGCCAACGG 23220

GCTATGTTGA CATTGATTGA AAAAGGCATG ACCCGTGAGC AAGCCTATGA CTTGGTGCAA 23280

CAAAAACAGC CTACTCTTGG GACAACCAAG TAGACTTTAA ACCACTTCTT GAGGCAGATT 23340

CAGAAGTAAC ATCACGTCTC ACACAAGAAG AAATCGATGA AATCTTCAAC CCAGTTTATT 23400

ACACCAAACG AGTGGATGAT ATCTTTGAAC GTCTTGGACT AGGTGATTAA TTAAAAAATA 23460

AACAGCGAGC TTCAATCTCG CTGTTTATTT TTTATCGAAA AGACTTAGTC TTCTTTTCTT 23520

TTAGTGAGTC CATAGGCTGC TAGTGTGGAC ATGAGTCCTG CGACTACTAG TCCTGCAGAA 23580

TCGTGAGTTC CTGTTTCAGG AAGTTTTTTC TCTGTTACCA CAGGAGCTGG ATCTTGAGGA 23640

AGAACTTTGC TTTCCTCAGC AGGAGCAGTT GATGGAGCTG GTTGGCTTGG GATTTCTAGT 23700

TTTGGTTTTT CTTCAGCAAT AGCGGCTTGT CCGTTTTCAT CGCCTACATG TGTTACCATA 23760

GTTCCGACTT CGACTATTTG AGTAACGGCT TCCTGTGCTA CGACACTATT TACAAGTGTT 23820

TTCACTTCCT TACCATCGGC AGAAGTGCTC ACAGAGTAGA AGTTGCTACG ATGTCCATTG 23880

ACGCCCTTAG TAATGACTTG TGTTTTTCCT TTGAGTAAGA GTGGATTTTC ACAAGTCACT 23940

GTGGTAAATG GAATTTCTTC TTCTTGGATA TCCAGTCTAG GTTTTACCTC AGTAGTTGGT 24000

GCAAGACCAC TTTCATCACC CTTGTGAGTT ACAGGAGCGC CAACTTCAAC CACTTGGTTT 24060

ATAACTTCTT TGGTTACCTG GCTATCAAGG ACTGTTTCTG TTGTTTTTCC ATTTTCAGTG 24120

AGTACAGAGA TGTAATGAGT TCGTTCACCT TTGACTCCTG CTGTGATAAT ATTTTCCTGA 24180

CCGGCTGGGA GGTTAGGATT TTCTTTCTTG ATAACTTCAA ATGGAATTTC TTCAGTTCTT 24240

GTGATGAGTT CTGGTCTGGT TTCAACATTG GCAGCCACTT CATTTTCATC TAGGCTTCCT 24300

GAATGAGTTA CAGCTGGTTT GAGGCCTTGA AGAGCGGCTT TTAGGTTGGC TACAAGCGTG 24360

TCAAGCTCAG CTTGTTTATT ACGGTTGAGG TTGTAATTTA GAGCTGTTTT AGCTGCGTCA 24420

AGGGCCTCAA GACTTTCTTT ACTATATCCT TCTAAGTTTG TAGGAATTTT AGCTAATTCT 24480

TCGCGGAGAG CATTATAATT AGCACGAAAG TAGTCTTTGT TGTGGTCTGC AAAGGCAGTC 24540

ATGAGTTCAA AGATTTCCTC TTCCTTGTAT TCAGCGCTTG GTCTATCTGC CCAGATTGAA 24600

AGCATACTTC CGACTGTTGG AAGATCTACT TCAGGATATT TGGTAGAAGC TAGTTGATTG 24660

AATGGTGTTT TTCCAGTATT CTCAATAGCT TTCTTGAGGA AACCACCACC ATCTTCTGGT 24720

TTTTGACCAA GAATGTAGTA CCAGTCACCG TTGGTATTCA AGAATTTATA GCCTTTGCTT 24780

GCTAGGTATT GAGGTGATGC GAGGTTATAT CCCCACCAGC CTTTAGACCA GTAAGAAATC 24840

AAGACATCTT TGTCAAACTG AACATCGTCC TTGTCTTCAT AGTAGAAGCC ATCGTTGAAG 24900

GCCATTGGTT GAAGCCCTCT TTCTTTGGCC ATAGCTGCGA GGGTGTTGGC ATATTCGGCA 24960

AATTTGCCAT AGAGTTGATA CCACTTGAGG TAGTACCAGC CTTGGGCACT AGTCGCATCG 25020

TTGGCGTATT CGTCAGTACC AAAGTTGAAA ATCTTTGTTT TACCTGCAAA GAAGTCCATG 25080

TATTTACCGA TGAGGGCTTT TACAAAGTTC ATCGCTTCTT CGTTTTTCAA GTCCATAGTT 25140

GTTTTTGAAA CTTTATCAAA GTGGGCTTGA GGATTTTTAA TACCTAATTT TTCCATGGCA 25200

ACCAGCATAG CATCCATGTG ACCTGGACTG TTAATAGCTG GGATGAGACC GATGTCCTTA 25260

GATTTAGCGT ATTCAATTAG CTCTGTTACT TCTGCCTGTG TTAGTGCAGT ACCGTTTGGA 25320

TCGTCGTAGT AAGCTTTAGT TCCTTCGATA ATAGCTTTTT TAACGTCATC ACTAGCATAG 25380

GTTTTTCCGT TGGCAGTAAT GGTCATATCA TCGAGTAGAA AGCGAAGTCC GTCATTTCCT 25440

AGAAGGAGAT GGACATCAGA ATATCCGAGC TCACTGGCCT TGTCTACGAT GCGTTTGAGC 25500

TGGTTCAGAG TAAAGTATTT GCGTCCAGCA TCGATTGAGA TTACCTTGTT TTTGGCAAGT 25560

TTTTCAACCT CACGTTTAGC TTCTTCTTCT TTTTGAGCTT CAGGCGTGAG GGTCAAGTTG 25620

TTGACAGTTT CTTGAAGTTT AGCAATGGCT TGATCAATCG TATCTTGTTG GGCACGGCTA 25680

AGGTTGCTAT CGAGAGAGCG AATAGCTTTT TCAGCTTCTT TTACGGCCGT GACGCTTTCT 25740

GCAGTATAAC GGTTCAGGTC TTTTGGTACC TCGTTAAGTG CTTGCTCTGC AGATTCATAA 25800

TCAGCTGCGA AGTATTCAGC GTTGGCATTT GCAAAATGAC GCATGAGTTT GAAGAGGCGT 25860

GATGGTGAAT AACGTGCAGA TGGAGTGTCA GCCCAAGCAG CTACCATACC ACCGATGATT 25920

GGGATATCAG CTCCTTCTGT TTTTGGTACA GAAGTGATTG GTGTGTTTTT AATACCATTG 25980

AGCCCCTGAT CGAGATTGTA CCAGCCTTGG CCATCAGCGT TTCGTCCAAG AACGTAGTAC 26040

CAAGCATCAT TGGTATTAAG GATTTGGTGA CCTTTTTCAG CTAGTAGTTT AGAAGAAGCG 26100

ACATCGTAGC CTCCCCAACC ACCAGTCCAC ATAGAAACGA TGATGTCTTT GTCAAAACTA 26160

CCAAAGCTTG TGTCGCTATT GTAGTAGATA CCGTCGTTAA AAGCCATTGG TTTGAGACCG 26220

TGCGATTTTA CAATACGAGC GAGGTCATTG GCGTAGGCAA TAAATTTTTC ATAGCCTTTT 26280

ACAGGGTAGC CTTCGTTTGG ATAGTATTTA TCAGCTTGAA GCACACTCCA ACCTTTAGCA 26340

TCTGTCGCAT CATTGGCATA TTCATCAAGT CCGATGTTGA AGATTTCAGT CTTTTTCGCG 26400

AAATAAGCAG CATACTTGTC GATAAGGGCT TTTGTAAAAG CGACAGCTTG TTCGTTGTCA 26460

AGATCGACAG TACGGGCTGA TTTCTTCCCA AAATAGCTAA AGTTAGGGTT TTGGATTCCC 26520

AATTCTTTCA TGGCATTGAG AATCGCATCC ATGTGTCCAG GACTATTTAC TGTCGGAATG 26580

AGACCGATAC CTTTATCTTT GGCATAGTTA ATCAGATCTG TCATTTGACT TTCTGTTAAG 26640

TGATTGCCGT TTGGATCGTT GTAATAATCA TTTGTACCTT TTTCAATGGC GCGTTTGACA 26700

TCGTCACTGG CATAGGTCTT GCCGTTAGCT GTGATGCTCA TATCGTCCAA CATGAAACGG 26760

AGTCCATCAT TTCCGACTAA TAGGTGTAAA TCAGTGTAGC CATAATGTTT CGCTTTATCG 26820

ATGATTTCCT TGAGCTGTTC TGGTGAGAAA TATTTACGTC CAGCATCAAT AGAAACAATT 26880

TTCTTTTTCG CTAGTTTTTC ATTTACAGTT GCAGCACGTT CCTTTCCTGC CTCTGTTGCC 26940

GGTTTGTCAG CCTCTGCTTT CGCTTCATCT TTTTTAGCTG GTTTATCCTT GTCAGTCTTG 27000

TCTGTATTTG ACTCTTTAGA ATCAACCTCT TTCGCTTCTT CCTTTTTAGG GCTAGCTTCT 27060

TCTGCCTTTT TATTAGCAGT TTCTTTTTCA GCAGAAGTTG GAGTTACCAC TTCTGCTTTA 27120

TCACTAGGAG TTGAACTAAC TTCCTCTTGT GGTTTTTCTT CTGTTTTTGG AAGACTAGCT 27180

ACCTTATCAG TAGCTGGAGT TTCTGTTTCT ACAGTTTTTG GAGCTTCTGG TTGAAGCACT 27240

GCTTTAGGTG TTTCCTCAGT CCGATTTTCG GATGATTGAG GGGAATCAGA AACCGTATGG 27300

ATGGTCGGTT GGTTTTCTGT AGTAGTAGGA GTAACTCCAT CGGCTGCAAC AGTCTGTGCT 27360

TGGAAGGCAA ATCCAATTAG AACAGAAGCT GCTCCTACAG CGTATTTACG AATAGAAAAA 27420

CGCTGTTGTT TTTCATGTTT CATTGCAAAA CCTCCTGATT GCATTGTTAT ATTGATAGCG 27480

ATTATATAAA TCAACGCCTT TATTTTATTT CTTATATTAA TTTCTTATAT TAACGAGAGT 27540

CAAGAGGAGA TGACAAAAAA CTATAATAAG TATAAAAAAA TATAAAATTT AAACTTAAGA 27600

TTTCAGATTG GTCGGAAAAA ATACGTATAT ATATCTAGTA TAATTTTTGG TTCTATTTCT 27660

ATAAAATATT CCACAAATTA TAGAATTTTC CAAAAATAGG TAAGCGCTAC CTTTTTGGTG 27720

TAGTATAATA AGCATAGAAA AAGCCCAAGC GATTAGCTCA GGTTTTCTTC TTAGTGATCA 27780

CGGTCACATG AGATAAATTT AATCTTGTAG TAATCAGATC GTTTGTAAGT TTCACTGTAT 27840

TCTAAAACTT GGCCAGTTGA TTCGAGTTTG GTGATTTTAG TTTGTAGGAC AGTAGGGAAT 27900

TGTTCATCGA CTCCGAGGAC TGAAGCTGCA TGTTCTGGAG TTGGAAAGAC TATTTCGTTG 27960

ATTTCTTCAA AGTGTTCATC ATTCATGTGA ATGTGGTAGT CTAACTTGAA ACGATTATAG 28020

ATAGAACTAT AGTATTCAAG GTTTGGATAA TTTGCGTTGA TATATTGTTC TGGGATGTAG 28080

GATGTATGGT AGATATAAAC GACACCGTTT GATTCGCGGA TACGTTCAAT CTTGTAGTAG 28140

AATTGATCGC CGCGTAGACC CAATTTTTCC AAGTAAACAA GCTTGTTTCC GCGTTCAATT 28200

GAAAGAACAG TTACCTTATC ATCTTTAGCA TTGAAGAGTT CAATATCTGA AAACTCTACA 28260

AGCTTGTGTT TGCGTGCACG TGAAACGAAG GTTCCTTTTC CTTGTTGGCG GACAATATAG 28320

CCATCTTTGG CAAGGTCGTT TAAGGCGCGA ACAACTGTGA TAGAGCTGAC ATCGTACATT 28380

GAAATGAGTT CTGCTTCAGT GTAAAATTTA TCTCCACTGC TAAACTGCCC AGAGATGATT 28440

TTATTTTTTA ATTCGTCTTT TATGTATTGA TGG 28473

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6749 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

CCTGATGGGT GGTATGCGAG GATACAGTTC TGAAAATCGC CGTTACTTAA TTAATGGACG 60

CGAAGTCACA CCTGAGGAAT TTGCTCACTA TCGTGCGACT GGTCAATTAC CAGGAAATGC 120

AGAAACTGAT GTGCAAATGC CACAACAGGC ATCAGGTATG AAACAAGGCG GTGTCCTTGC 180

AAAACTAGGT CGAAACTTAA CAGCAGAAGC GCGTGAGGGC AAGTTGGATC CTGTTATCGG 240

ACGAAACAAG GAAATTCAAG AAACATCTGA AATCCTCTCA CGCCGCACCA AGAACAATCC 300

TGTTTTGGTC GGAGATGCAG GTGTTGGTAA GACAGCAGTT GTCGAAGGTC TAGCGCAAGC 360

CATTGTGAAC GGAGATGTTC CTGCTGCTAT CAAGAACAAG GAAATTATTT CTATTGATAT 420

CTCAGGTCTT GAGGCTGGTA CTCAATACCG TGGTAGCTTT GAAGAAAATG TCCAAAACTT 480

AGTCAATGAA GTGAAAGAAG CAGGGAATAT TATCCTCTTC TTTGATGAAA TTCACCAAAT 540

TCTTGGTGCT GGTAGCACTG GTGGAGACAG TGGTTCTAAA GGACTTGCGG ATATTCTCAA 600

GCCAGCTCTC TCTCGTGGAG AATTGACAGT GATTGGGGCA ACAACTCAAG ACGAATACCG 660

TAACACCATC TTGAAGAATG CTGCTCTTGC TCGTCGTTTC AACGAAGTGA AGGTCAATGC 720

TCCTTCGGCA GAGAATACTT TTAAAATTCT TCAAGGAATT CGTGACCTCT ATCAACAACA 780

CCACAATGTC ATCTTGCCAG ACGAAGTCTT GAAAGCAGCG GTGGATTATT CTGTTCAATA 840

CATTCCTCAA CGTAGCTTGC CAGATAAGGC TATTGACCTT GTCGATGTAA CGGCTGCTCA 900

CTTGGCGGCT CAACATCCAG TAACAGATGT GCATGCTGTT GAACGAGAAA TCGAAACGGA 960

AAAAGACAAG CAAGAAAAAG CAGTTGAAGC AGAAGATTTT GAAGCAGCTC TAAACTATAA 1020

AACACGCATT GCAGAATTGG AAAGGAAAAT CGAAAACCAC ACAGAAGATA TGAAAGTGAC 1080

TGCAAGTGTC AACGATGTGG CTGAATCTGT GGAACGAATG ACAGGTATCC CAGTATCGCA 1140

AATGGAAGCT TCAGATATCG AACGTTTGAA AGATATGGCT CATCGCTTGC AAGACAAGGT 1200

GATTGGTCAA GATAAGGCCG TAGAAGTTGT AGCTCGTGCT ATCCGTCGTA ACCGTGCTGG 1260

TTTTGATGAA GGAAATCGCC CAATCGGCAA CTTCCTCTTT GTAGGGTCTA CTGGGGTTGG 1320

TAAGACGGAG CTTGCTAAGC AATTGGCACT CGATATGTTT GGAACCCAGG ATGCGATTAT 1380

CCGTTTAGAT ATGTCTGAAT ACAGTGACCG CACAGCTGTT TCTAAGCTAA TTGGTACAAC 1440

AGCAGGCTAT GTGGGTTATG ATGACAATAG CAATACCTTA ACAGAACGTG TTCGTCGCAA 1500

TCCATACTCT ATCATTCTCT TGGATGAAAT TGAAAAGGCT GACCCTCAAG TTATTACCCT 1560

TCTCCTCCAA GTTCTAGATG ATGGTCGTTT GACAGATGGT CAAGGAAATA CAGTAAACTT 1620

CAAGAACACT GTCATTATTG CGACCTCAAA TGCTGGATTT GGCTATGAAG CCAACTTGAC 1680

AGAAGATGCG GATAAACCAG AATTGATGGA CCGTTTGAAA CCCTTCTTCC GTCCAGAATT 1740

CCTCAACCGC TTTAATGCAG TCATCGAGTT CTCACACTTG ACTAAGGAAG ACCTTTCTAA 1800

GATTGTAGAT TTGATGTTGG CTGAAGTTAA CCAAACCTTG GCTAAGAAAG ACATTGACTT 1860

GGTAGTCAGT CAAGCGGCTA AAGATTATAT CACAGAAGAA GGTTACGACG AAGTCATGGG 1920

GGTTCGTCCT CTCCGTCGCG TGGTTGAACA AGAAATTCGT GATAAGGTGA CAGACTTCCA 1980

CTTGGATCAT TTAGATGCTA AACATCTGGA AGCAGATATG GAAGATGGCG TTTTGGTTAT 2040

TCGTGAGAAA GTCTAAGACA GAATTTTGAG GATAAAAAAG AAGGAGCCAG CTGAAAAAAA 2100

CTGGTTCCTT TTTAGGTACG ACAGGCATGT CGTATAGTAG AAGTGTATTA TTCTAGTTTC 2160

AATATACTAT AGTAGCTCAG AAGTCGGTAC TTAAACGTGC TATATCAAAA CCAGTCCTGG 2220

AAAAACGTGG ACTGGTTTCG TGTTTGGATT ATTACCTTGA ACGACATGCG TTAAAAGTTA 2280

GTTGAACCGC CGTATGCCGA ATGGTACGTA CGGTGGTGTG AGAGGGGCTA GAGATTATCC 2340

CCTACTCGAT TTTAAATCAC ATGACGTTCA AAGGCATCAT CTGAAATCCC TTGTTCCAAG 2400

ATGAGTTTTG CCCATTCTTT AGCAGAGAAG AGGCTGTGGT CCTTGTAGTT TCCGCAAGAT 2460

TCGATGGTTG TCCCTGGGAC ATCTTCCCAA GTAGTAGTTT CAGCGATTTC CTTGAGCGAA 2520

TCCTTGATAA CAGCTGCGAT TTTAGCACTG GTGTGACGTC CCCACATAAT CATGTGGAAG 2580

CCTGTGCGGC AACCAAATGG TGAACAGTCA ATCATGCCGT CAATGCGGGT ACGGATGAGT 2640

TTGGCTAAGA GGTGCTCGAT AGTGTGAAGG CCGGCAGTAG GGATAGAGTC TTCGTTTGGT 2700

TGCACCAAGC GAATATCATA ATTGGAGATG ATGTCTCCTT TTGGTCCTGT TTCTTCCCCA 2760

ATCAAGCGAA CATAGGGTGC TTTGACAATG GTGTGGTCAA GTTCAAAACT TTCGACAATA 2820

ACTTCTTTTG ACATGGTAAA TCCTTTCAGT TTTCTTCTCT CATTATATCA TAAAGGTTGC 2880

TCCTGAGACA GAGAGAAAAC CTCTCCGAGG CTGGAGAGGT TGAAATCTTT ACTTACGATA 2940

TAAGCGGTCG TATTGGTAGT ATGGGTCAAA GGTTACGTTG ATACCCAGTT TACGAAGGAC 3000

ATTCTTGTCT TCATCAGTCA AGATGATGGT TGAGTGGGCT TCGCTTCCTT TGAGGTTGCC 3060

GAGTTCTTCC ATAGCGCGGG CAGCATCAGG ATTTTCTGTA GCTGTGATAG CAAGTGCAAT 3120

CAGGATTTCA TTTGAATGAA GGCGTGGATT GCGGCTACCG AGATGATCGA TTTTAAGACC 3180

TTGGATTGGC TTAACAACTT CAGGCTCGAT TAGTTTTACT TCTTTAGCGA TGTCAGCTGA 3240

TTTTTTGATG GCGTTGATCA AGGCAGCGGC TGTAGGACCA AAGAGTTCTG AGTTCTTACC 3300

AGTGATGATT TCCCCATTTG GCAATTCAAA GGCTAGGGCT GGTCCACCAG TTTCTTCTGC 3360

TTTTTGGCGC GCAACGACAG CAACCTTACG GTCTGCAGGT GTGATACCGA GGTCGTTCAT 3420

GAGCAACTCA ATTTTCTTGA CGGCAGCTTC GCCAACTTTT TCAGCTTTGA AGTCAAGAAC 3480

TGTTTGATAG TAACGGCGGA TGATTTCTTG TTTAGAAGCT TCGACAGCGG CCTCGTCATC 3540

TGTAATAGCG AAACCAACCA TGTTGACACC CATATCTGTC GGTGAAGCGT ATGGTGATTT 3600

TCCGAGAATA CGTTCCAACA TGCGTTTGAG CACTGGGAAG ATTTCGATAT CACGGTTGTA 3660

GTTGACAGTG GTTTCTCCAT AGGTTTGAAG ATGGAAGGGG TCAATCATGT TGACATCATC 3720

AAGGTCAGCT GTGGCAGCTT CATAAGCCAA GTTAACTGGA TGATGAAGGG GAAGATTCCA 3780

AACAGGGAAG GTTTCAAATT TAGCGTAGCC AGATTTGATG CCATTGATTT GGTCGTGGTA 3840

CATATTGGAC ATACACGTTG CCAATTTTCC AGAACCAGGT CCAGGAGCGG TTACGACAAT 3900

CAAGTTGCGA CTGGTTTTGA TGTAGTCGTT TTTGCCCATG CCTTCTGGGG AAATGATGTG 3960

ATCCATATCC GTCGGATATC CTTTGATTGG ATAATGAAGA TAAGAATCAA TTCCGTTTTT 4020

CTCAAGTTGA TTGCGGAAGG CATCTGCAGC GGGTTGGCCA GCGTATTGTG TAATGACAAC 4080

GGAACCAACA AAAATCCCTA ATTCATTGAA TTTATCAATC AAACGAAGAA CTTCTTGGTC 4140

ATAAGAAATG CCTAAGTCGC CACGTGCTTT GGAATGTTCA ATGTTGCTAG CATTAATGGC 4200

AATCACAACC TCAACCTGCT CTTTCAATTC TTGCAAGAGC TTGATTTTGT TGTCAGGTTC 4260

ATAACCAGGA AGGACACGAG CAGCGTGGAA ATCTTCTAAC ATTTTACCGC CAAACTCTAA 4320

GTAGAGCTTG CCGTCAAATT GGTTAATGCG CTCCAAAATA TGGTCGCGTT GTAAATTCAA 4380

ATATTGTTCA GAACTAAAAG CTTGTTTTTT CATTTTTTTA CCTCTGGACT CTATTATAAT 4440

AAAAAATTGG AAGTTAGGAA ACTACGGAGC TAAAAAAGAA ATTAAAAAGA TTAAGCAAAC 4500

GCTTGCACAA AATTTTAAAA AGTGCTATCA TAGACTATAG ATTATGAAAA TAATGAGGTA 4560

AACAGATGCA AGAAAAATGG TGGCACAATG CCGTAGTCTA TCAAGTCTAT CCAAAGAGTT 4620

TTATGGATAG TAATGGAGAT GGAGTTGGTG ATTTGCCAGG TATTACCAGT AAGTTGGACT 4680

ATCTAGCTAA GCTAGGAATC ACAGCAATTT GGCTTTCTCC CGTTTATGAC AGCCCTATGG 4740

ATGATAATGG CTATGATATT GCTGATTATC AAGCGATTGC GGCTATTTTT GGAACCATGG 4800

680

AGGACATGGA TCAGCTGATT GCAGAAGCTA AGAAGCGTGA CATTCGTATC ATCATGGACT 4860

TGGTGGTCAA TCATACCTCA GATGAACATG CTTGGTTTGT CGAAGCCTGT GAAAATACTG 4920

ACAGCCCTGA GCGAGACTAC TATATCTGGC GCGATGAACC CAATGACCTA GATTCTATCT 4980

TTAGTGGGTC TGCTTGGGAA TACGATGAAA AGTCAGGTCA ATACTATCTC CACTTTTTCA 5040

GCAAGAAACA GCCGGATCTC AACTGGGAAA ATGAAAAACT TCGCCAGAAA ATTTATGAGA 5100

TGATGAACTT CTGGATTGAT AAAGGTATTG GTGGTTTCCG TATGGATGTT ATTGACATGA 5160

TTGGCAAAAT TCCTGACGAG AAGGTAGTCA ATAATGGTCC TATGCTCCAT CCCTATCTCA 5220

AGGAAATGAA TCAGGCGACC TTTGGAGATA AGGATCTCTT GACAGTAGGG GAGACTTGGG 5280

GAGCAACTCC AGAGATTGCC AAGTTCTACT CTGATCCAAA GGGGCAAGAA TTGTCTATGG 5340

TCTTCCAGTT TGAACATATC GGTCTTCAGT ATCAGGAAGG TCAGCCTAAA TGGCACTATC 5400

AAAAAGAGCT GAATATCGCT AAGTTAAAAG AAATCTTCAA CAAATGGCAG ACAGAGTTAG 5460

GAGTTGAGGA CGGCTGGAAT TCCCTCTTCT GGAACAACCA TGACCTCCCT CGTATTGTCT 5520

CAATCTGGGG AAATGACCAA GAATACCGCG AAAAATCTGC CAAAGCCTTT GCAATCTTAC 5580

TTCATCTCAT GAGAGGAACT CCTTATATCT ACCAAGGTGA GGAGATTGGG ATGACCAACT 5640

ATCCGTTTGA AACACTGGAT CAAGTAGAAG ATATTGAATC TCTCAACTAT GCGCGTGAGG 5700

CTCTTGAAAA AGGTGTTCCG ATTGAAGAAA TCATGGACAG TATCCGTGTT ATTGGACGTG 5760

ACAATGCCCG TACCCCTATG CAATGGGACG AGAGCAAAAA CGCTGGTTTC TCAACAGGTC 5820

AACCTTGGTT GGCGGTTAAT CCAAATTACG AGATGATCAA TGTCCAAGAA GCGCTGGCAA 5880

ATCCAGATTC TATTTTCTAT ACCTATCAGA AACTGGTCCA AATTCGCAAG GAGAATAGCT 5940

GGCTAGTTCG AGCTGACTTT GAATTGCTTG ATACGGCTGA TAAGGTCTTT GCTTATATAC 6000

GTAAGGATGG CGACCGTCGC TTCCTAGTTG TGGCTAACTT GTCCAATGAA GAGCAAGACT 6060

TGACAGTAGA AGGAAAAGTC AAATCTGTCT TGATTGAAAA CACTGCGGCT AAAGAAGTAC 6120

TTGAAAAACA GGTCTTGGCT CCATGGGATG CTTTCTGTGT GGAATTACTA TAAATATTTT 6180

TTGCAGAAAA ATTTAAAATT GAAATCGTAT AAAAACAAGG GAGGACTGTA TAAAAGACAG 6240

AAATCCTTTG TTTTTTATAA CCAAAGTTTA TAAACTTTCA TTCTTGAAAT TCAATTAACT 6300

TTACAAATTC CCACTATTAA GGAGAAAGAA GATGAACATA AAGAAGCGTG TCCTTAGTGC 6360

AGGCCTGACT TTTGCATCTG CTTTGCTTTT ACCCAAATCA TTCATACCTC TCTCAACTAG 6420

ATGTAACTTA CAAAACCCCT GACCTCATGA GCCACTTTCT TCCTCCTCAT GAGGTCAGTT 6480

TTACTTTCTG CTGTTCCAGT ATCGTTTTTC CTCGCTAGAT TTCCTCAAAA GGGCAGACTC 6540

CTCCCTTGGT GCGTCACACG ATTTTTTCAT CTCGACTGTT CTTTAATGCA TCATTAACGA 6600

CGCTTTTCTT CTAGGTGGTT CATAAGGAAC AGGAAGATTC AGGTTGACTT TTCTAATCCT 6660

AGAATAAAGT GCTGAAAACA ATTCGGAATA GGCATAGAGA CTAGACAATT TGAGGAGCTG 6720

CTTGCGTCCT GTTCGAACAC ATTTTCCGG 6749

(2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1842 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

TCTACCCATG GACTTTGAGG CATTCATTGT TCCATCTTCT AGTGGCGAAT CTTTTGATAC 60

AAACGATTCA ATTCACTTGG ATAGTGAAAC TCTCCCGCAA ACATTTTTCT GGTTAACTCA 120

ATCCAGCTGA TATTTCTTTC AGCCAAAATA ATGGACAAGT TCTCCCAAAA TCGTTCAGCC 180

ATATTGCTTC TCCTTTAGTT AGATAAATAA TGTGTTTGCG CCATGTAAAT CAATTGTTTC 240

GTATCTCTTG GCAATAGAGC TCTAGCCTCT TCCAAATTCA GACTTGGATA AACTCGCTTA 300

TTTGAAACCG CAAGAGGAAG TCTGATGGTT AGTTCAGGAT TTTTTAAAAT TATCTCAACG 360

AAATCCGTTA ATCTTAGATT GTCACGGTTC TTAAATCGTA ATAAATTGGG AGATAAAAAC 420

TCAAAACAAT CTGAAGAATA GCTCATCATC TCAATTAATT TGTCCTTTGT CATTTCAGAA 480

ACTGAATGAC AAGATACCTC TATGCCATAG TTTTGGAAGA AATCTAAAAG AAGTTGATTT 540

CTTTGTCTAT TTTTACTTAG ATAGAGATCA ATCATGGGAG ACCTCCCAAA GATTCGGTTC 600

CATTTGATAT TCTGACACGA TTAAGGAATC TAATAAATTA AGGAATCTAA TAAATTTGCG 660

AAGTTAATCG GTTTCTTGTC TTCATCATAA GCTTTTACAG TTACTTGGGT TGTAAGTATT 720

CCCTCTTTTC CCTCGGCTCG ATAGCCTTGT CCATATAAAA CAAAAACGAG ATTTTGATGA 780

TCATCTACAA AGGCATCAAC CCCATTCTTT ATGTCTTGAC TTTCAAGGAA TTCCATAACG 840

TTTTGAAGAT AGGATTCGTA AAATAGTGGG TAGTTATGTT TTTTATGGTA ATCATCTAAA 900

AATGTCACTT CAAACTCACA TGGAGAGTAA TTTTGACTTT GAACAGCCTA AAAGTGCCAT 960

CAAATTTGAA TTGGAATAAA TCAAATAAAT AGCCCCATCC TCATCAATCC AACCTTTGCT 1020

CAAAGACAAC TCCAACCGAT CTTTTAAAAC TGAGTAAACC ACCTTAACCT CCAGTTTCAT 1080

ATTCTTATAC CGTTCACTCT CAAATAAAAG TTTGGGGAGC TTATAATAAC GCTCTGATGT 1140

CTGATATTGA TTAGCGGTAA TACGCTTCAT TATTGTCCCT CCAAGACTAA AATTCCAACA 1200

TTTCCAAATT CATCAAATCG GATTAAACCT ACTTGTTCCA TTTCATCAAC TAACTGAGTT 1260

GCTTTTACCC AAATCATTCA TACCTCTCTC AACTAGATGT AACTTACAAA ACCCCTGACC 1320

TCATGAGCCA CTTTCTTCCT CCTCATGAGG TCAGTTTTAC TTTCTGCTGT TCCAGTATCG 1380

TTTTTCCTCG CTAGATTTCC TCAAAAGGGC AGACTCCTCC CTTGGTGCGT CACACGATTT 1440

TTTCATCTCG ACTGTTCTTT AATGCATCAT TAACGACGCT TTTCTTCTAG GTGGTTCATA 1500

AGGAACAGGA AGATTCAGGT TGACTTTTCT AATCCTAGAA TAAAGTGCTG AAAACAATTC 1560

GGAATAGGCA TAGAGACTAG ACAATTTGAG GAGCTGCTTG CGTCCTGTTC GAACACATTT 1620

TCCCACCACG TGAAGAAAAA GATGGCGGAA GCGTTTGATT GTTAAAGTTT GGAAGTCACC 1680

TCCAGCTAGA TGTTTGAGAA AAAGATAGAG ATTGTAGGCG ATACAGCTCA TCATCATACG 1740

AACTTCGTTT TTGATTAAGG TTGAACTATC CGTTTTATCG CCAAAAAATC CCTCCTTCAT 1800

CTCCTTGATG AAATTCTCGG CTTGACCACG TCCACGATAA AG 1842

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19390 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

TCATCTTTAT CTCCTCGAAA TTTTCTAATA TAGCCATTAT AACAGAATTT TGTGAAAATT 60

CCTATTATAG TAAATCACTA TTTCAGTATA AAAAGAAAAA ACGAATCAGA CGATTCGCTC 120

TTCTTAAAAT CTGAAAATAG CTTTCCAGAA AGGATTAGCC GATTTTTTGC AGATTGAGCA 180

CTGCATCGTG ACTCATCAAG ACTTGACCAT ACTCTTGTAA GACTGAGCGA CTGATATCAC 240

TATCGTCTGC AAACTCGCGC ATACGGGCCA ACAGCCAAGC TGGATATGGG CTTGGATGAT 300

TTTCAATATC CACTAAAATG GTCAAATAAT AGCGCTCGTT CATTTTGTAG AGTTCAGAAG 360

TTTCCATTTC AAAAGTCACT GTCTTGGCAA AAGCTACCAA GTCAGCCAAC TTAGCAAAAG 420

AAAGGATGTA GTAGATGTAA GGTTCTTTCT TACTCTCAGC TTCTTGTTCA GCCTGCTCTT 480

GCTCTTCTTC CTTGACTTCA ACTTGCTCAA GAGATTGAAT GGCTTCGATA TCATCCTTGG 540

TTTTGTCTGC GATGCTTTTT TCCAGGGTTT TGATAAATTC ATCTGGAGAC ATTTGAGCCA 600

ATTCTTCCAT ATCTGGCAAA TCCGATAAGT CTTCAAAATC TAGATTTTGG TCAATCTTTG 660

ACTTGGTCAC AAAGACATCT ACCTTATCAG GTTTTGGAGT CACACGGAAG CTCAACATGC 720

CTGTATCCAG AAAGCTATCA GGCATCTCTA GCTCATCCAA GATAGCATAA AAGAACTCTT 780

CTGTTTTTTC TTGAGGAACG AGAAAGTCAG CAATCTCCAT TCCACGATCC ATCAAATCCT 840

CTAAAGATAT CGTGATTTTT AAAGTTGTAT CACTAATTTG TTTCATTTTC ATTGCTAGTA 900

ACCTCATACT TTCAGTTCTA TCTATTATAC TAGATTTTTA CGATTTTATC AAAAGAAGGC 960

TCCTCTATAC GGATAGATTT TCCCTAGGGT CTTTCTATAG GAGACTCCAA AAGAAAATTT 1020

CTGCAGACAG ATAGAAAAAG CCTTCAAAAT CGGCTAAGAG CCGACTTTGA AGACCTTATA 1080

CATCAGAATA CTTATAATTT AAAGGTTGCT ACACCGAGGA TAGAACGATT TAAGTTTCTG 1140

AGAATTTGAA GACTTTGCTC AAATTTCTTA TAACGAGTCA CTCCGTACTC TTCAACAAGA 1200

AGGACTGTAT CTCTTTCCAA AAGAGATGAT ACATCCTGTA AATCTACAAA ATGCATTCCT 1260

TTTAAAGCTT CTTGACTCTG TTTCAATTTA TCTAAGATAG CTTTATTTGA GCTAACGATG 1320

GTCAATTCCT GTCCAGTATT TTTGTATGAC AAAACATCTG CTAGGTTAGC AATTGTTGTA 1380

ATCTCTGTTA CAAAATCAAT TTGATACTGA GAAAAATCAC CTACTCTATT GATTGTTGGA 1440

TTAAAGAGAT AAACTAACAC ATTTCCCATC ACAACCAAAA TCACACAAAC CACTCCAATA 1500

ACAACTAAAC GAAGAATCAG ATTTTTCACA TTTAAGCCAA GCGCTGTTTC ACCATTTGCG 1560

TTCAATTCTT TAGAGTTGAT GGTTTCCAGT TTTTCAATTT TCACATTTGC ATAGGCATGT 1620

TTAAATTTCT CAATCAACCC ATCAATTTTT TTCTCTAACA AGTTATTGGC ATCTTTACTT 1680

GATGTCAAAA TTTTCACACC AACCCCTGCA TCGTCAATCA TATAGTAGAC GGTCAATTTT 1740

TTCCACCAAT AGTCATTCGT TGAATTTTTC AAGGTTGTTT CTGTCGTGTC TAATTCACTG 1800

GCAATTTTTT TCAACTCACT GGGTTCTACA TCATTGAAAA GATAAGCTCC ATTCAAATTA 1860

CCATCAATCA ATTTCCCATA AAAATCACTA TAACCACCAA TTTGATGATT CAAAATCGTT 1920

TTGTCCGACT CTTTTGGAGG AGTGATTTTA TAGATAAGAT AAGTTGAATA ACTTGTTGTA 1980

TCTTTGACAG TGTTTTTATT CCTAACTGCT TTAATTGTAA ATGGTACAGC AATGAGAGCA 2040

AATAAAGCGA TGAGAGCTAA AATATTTGCT TTTCGCTTTT TATAAAGATT TGCAAACAAA 2100

TCAGCTACTG AATAATGTTC AAACATGATT TTTTTCTCCT TTGTTTAGTA GATACTAGTT 2160

TTCCTTTGTA AGCATTTTTG CTACAAATAT AATCACAAGA ACAATTCCCC AGAATTGCAT 2220

TGTAAATAAA TTGAAGAAAC TTTCTGAAAA GCTGCTTCTT GGCATAAAGA ATAGATTATT 2280

CAAGATGAGT AGGGATAAAG CAAATAGGAT TGTCCTTGAG CGATAGGCTA CTTGCAGCAT 2340

GGCTATAAAT AATACGCCGA GTAAGAAACT AAGCAGAAAG ACTCCAATCA TACCATAGTC 2400

GGTATACAAC TCCATGATAT AACTACTTCC GATACCATGC CCTTTCAAGT ATTCCTTGTT 2460

CAAGACAAGA TAGGATAGAT TGTGGGCATA ACTATTACTA TCAATAGCTA GTTCCACACT 2520

ATTGGTTGTA TGTTCAAAGG CTTTTCCTCC GAAAATGGCT CCCAAACTCC CCCTTGCAAA 2580

ATAATCAAGA ACAGGACCAA AAGTAAAATT ACGGAAATCT CGGTAAGGGA GGCTACTGTT 2640

AAATAGAAAA CCTCGAGCCA GAACACCAAA ACTAGTCCCT TGTTTATAGA TAAAGTCAAG 2700

TAAGATATCC CAGAAACCTG TATGGGAAAC TTGGACATTA TCCCGTACAT AATTGAGTAC 2760

TCCCATCGCT AACATGAGAA TAGGAGAACC TACAAAAATC GCTAACTTTT CTTTAAACCC 2820

AATCCATTTT CCTTTTTCAG TTTGCTCCCG CATAAAGTAA TAAACAAAAG CAAATmAAAT 2880

ACTTAAAATA AAGGGATTTC GTGTCCCAAT TGCCAAATGA ATAGTATTAG CTGCAATAAA 2940

GGAGACAAGC ACTGCTGTGG CCTGCAATTT CTTTGGCTTG GTTGCCAGAT ACATACACAT 3000

TGCATAGACC GTAAAGGTAG ACAAAATGTA GGTAAAATAA GGCAGTTTAC TTTCAAAATT 3060

TGCATAGTAG GCATAGTAGG AAGTCTGCAA ACGATACAAG AGCCGTTCAA ATAACCGAAT 3120

GAAATAGAAA GGATAAGTTA GAAGAAAAAC TCCTAGTGAT ACAAAGCGTA ACCGCTTGAT 3180

ATAAACCTCT TTTAGAGAAT TTCCTATATT TGCTACTTTT ATTTTCTTCC TAGCTATGAA 3240

GTAACGAGCC AGAATGCCTC CTGTGGTCAA GCCCAGAATC GAAATCATGA CAACTATAAA 3300

GGCAAAACGA TAGGCTATTG GATGATAGGT ATCCAAAGCA CCATCCCTAA AATAATCAAT 3360

GGTCGGTCTT GATACCAGAA ATACAAAAAT GGTTAAATAG AAAATAAAAT GGATTAAGTA 3420

ATACTTGATA TCATTCCAAC AAGCAATTAA GCTACTAACC AACAAGAACA ATAAAGTAGA 3480

AAGTAAGCTA ACATTATTAT TATTAAACAG ATACACAATT CCACTTACTA GCGTCAAGGC 3540

ATAACTGACT ATGGTCAAAC TAAATAATAA TCGTTTCCCA TCAATCACTT GGTCACCCCC 3600

GTTCTAATGT AATTTTTTAG ATTTTTCAAT ATTTTTCAGT AATAAGAATC GATATAAGGA 3660

AATATTTATG AATAGGGCCA AAGCACTAAT TCTTCTCCCC TTACGGAAAA TTGGATTCCT 3720

AGAAATAGCA AAGGCATGGC CTTTTAAAAA ACGATGAATC TGAGAATAGG CTTCAAACTG 3780

TTTATACTGA TCATCTAGCA ACATCTTATC CAGAATAAAG AAGTGGGCAT AGGCCAATCT 3840

GAAAAAAGCG ACCTCTTTCA AGTCAGGATA GTTTTTCACA ACTTCATTAT AAAACTTTTG 3900

GTAGATATCA ATATAGGCTA AATCCTTCTC TGCATAGGGT TTGGTCGTAA TACTATCCCC 3960

TCTATGGAAA TAGTAATAAT AGGGTTTAGT ATTAACCACA TACTTCTTGG CCAACTTGAT 4020

TAAATCAAAA TGGTAATAGG CATCTTCGTA AATCAACCCC TTAGGAAAGG ATAGGGCAGT 4080

TGCAATCTGT CTCTTGATTA GCTTATTGCA AATCGTCCCA GGTATTTTTT CACCTATGAG 4140

GTATTCCTTT AGAAATGTTT GAGAATCACA GACAAAATAG TCATCCTGAT TGGCTGACTG 4200

TGGGCTTTCA TCATTAGCAT AGACATTCAT GACACCACAG CTCGAAACAT CCGCATCTTC 4260

TTGAACTAAT TGCTCATATA AGCTCTGAAT CATTTCTGGA TGGATATAAT CATCTGAGTC 4320

AATAAAAATC AGATAATCCC CGTGAGCCTG CTTCATCCCA TCATTTCGTG CTTGCGACAA 4380

TCCTTCGTTC TTTTTATGAA GCACTGACAC CCTGTCATCT TGTTCAGCGA TTGAATCACA 4440

CAAGCGACCA CTTTCATCTG TTGCACCATC ATCAACAAGA ATAATTTCCA GATTTTGATA 4500

GGTCTGCTTC TGAATGGAAG CTATCGATTT TTCTAGGTAC TGCGCCACAT TATAGACTGG 4560

CACAATCACA CTAATTAATG CAGTTTCCAT GCTACTCCTC TAATAGTTTT TCTACTTGTT 4620

CGATTTGTTT TGTAATTGTA AATTGTTGAA TGAATTGGCT AGCCTCATCG ACATCAAAGT 4680

TTGAGGCAGA AGTCATGTAA TTAGTAATCG CCTGAGCTGC CTCTTGATTG CTCTCAATGA 4740

TTTGTCCAAA TCGTCCTTCT TGGGATAATT CCTCAGCCCC TCCAACGTCC GTAGAGATAA 4800

AAGGGAGTCC CAGACTCAAG GCCTCCACAT ACACTCCAGG AAAACCTTCT TGTTTAGACA 4860

TAGACAAAAG AACTTTCGTC TGAGATAGAT ACTGATAAGG ATTTTTTTGA TAACCAAGGA 4920

AATGTACATA GTCCTCAATC CCATACTCTT TGACTCGTTT TTTCAGTTCC TCTTCCATAT 4980

CACCAGCCCC GATAAAATAG AGATGATAGT TTTTTCCCTC TTGGTGTAAT AATCGTATCA 5040

CTTCCACTAC ACGGTCAGAA CCCTTATTTT CCTCAATCCG TCCGATAGTA CAGATACTTT 5100

GAGGAGCAAT CTCGATATCG ATCTTCTCTT GAGATTTTTC TAGAATAGTC TGAAAATCAT 5160

ATCCATTGTA GATTGTCTGT AATTTAGAAG TATAATCTGG ATAAACTTCC TTGATAGAAT 5220

TGCTGGTCTT TTTTGAAATC CCTACAATTG TATTCGCAGC ATCCAACTGG CTTCTATGTG 5280

ATTCTCTTTT AGAGCTATCC TTAAGAAGTT CTTCAATACT TCCATGAATC CAAGATATCT 5340

TCTTGACTTC TCTTCTTTTA GAGAACAACA GTGGTGGATT CATAATGGTA AAAGAAACTT 5400

CAACATCATA ATCATCTTTT ACAAGCAAAC GACGAGTCAG TCTTGGAAAA TAAATTCTCA 5460

TTCTCCACAA AAAAGCTCGT AACCATCTGG TTTGGCGATA ATCTTGAAGG GATTTTAAAA 5520

TGCGTACATG CTTTGGAACA GATTCATATC CCTTGTCAAA GTGCTCCATT TCAAGAATAT 5580

CAATATCATA CTTTTCTGGA TCCAGATTTG AAACAATGGT TGATAGAATC TTCTCTGCAC 5640

CACCTCCAAG AGAAAAAGAC CACATAAAAA ATAAGATTTT TTTCTTAGCC ACCATATTCT 5700

CCCTTGTATT CTGTATAAGA CTTATCCATA TCAGCGATGA CAGCATCATG ATGCGGTACC 5760

TGCTTGTCTG CTGGTGGAGG CGTCATATAA TCCCCAAAAG CAGTTCTGAG ATAGACATCA 5820

TAGCCGATTG GAATAGGCAT CTCTGTTCCT TCAAATGGCA AGAAAAGATT GTCTTCAAAA 5880

GATGTGATTG GGTACTTGTT TCTCATGTAG CCAGGACCTG AGCATAATTC TGTAATGCCA 5940

TCACAATCAG CCAAATCATA CTTAGTCATT TCTTTCTCAG CTTTTTTCCA GATGCGATAA 6000

CGGAGAGATT TTGGAGTCAA ACCCAGTAAA ATGCGACTTC CCCATTTCAT GAGATCACCA 6060

TGCTTTTCTG GAATAGTTTG CGCACAAAAG AGTGAATAAA TCAAGGCCCA ACGAACCTGT 6120

TTTTTCCGCT CAGCTGGATT TTTCGGATAA TAATCCAAAG GCAAAACATC CAAGGCCAGA 6180

CCATGTGGCA AATCCAAATC CTGCTGATAA GGCTTGATAC AGGTGGTTTT CTTGTCACGA 6240

ATGGTAATAA AAAGATTACG ATCAACAAAA TCCTTGTGAC TCTTTGACAA GAAATAACGT 6300

TCATCTGCAT AACGAGGCCA TAATTCTGCT AATTTCTCAT AATCTTTACG AGGCATAAAA 6360

AAGTCTAGGT CGTCGTCCCA AGGAATAAAT CCCTTGTTTC GAAGGGCACC AATAGCGCCT 6420

CCGCCACAGA GATAACAGAG CAAATCATGT TCTTTACAAA AGGCCACAAA ATATTCAGCC 6480

ATCTCCAGAC TACGAGCCTG AATTGCTTTT AAATCAGTCA TATTGTTCAT TATTCTTTCT 6540

ATCGTATCGT TTCATTATAC CACAAACAAG GGGTGAAAAT CTATTGCAGA CTGTAAAAAA 6600

TCAAAGCCTG ACTGCTATCC AAATAGCTAT CAAACTTTGA TTTTTCTGTC TTATACTCTT 6660

CGAAAATCTC TTCAAACCAC GTCAGCTTCA CCTTGCCGTA GGTATAGGTA ACTGACTTCG 6720

TCAGTCTTAT CTACAACCTC AAAACTGTGT TTTTAGCAGC CTGCGGCTAG CTTCCTAGTT 6780

TGCACTTTGA TTTTCATTGA GTATTATCTT ATCTTAAGCC CATTTGAGCG AGCTTGGTTT 6840

GATATTTGTT TTGATCAACC AGCAGGCCCA AGCCCCCATA AACATCATAG GCATCTACCC 6900

AGTCACCCAG TTCTGGAATC GTCAATTTTT CAATACCATT TTTTGCTCCA TCCAAAACAG 6960

ATAAACCGTT TGTTAGGAGG AAAGTATAGG GTACGTTGGT TGAGGTCATA GCAAAAACCT 7020

TTCCAAGAGC TTCAGAACCA GTGAAAAGTT TAGTGGGATC TTTAATTTGC TCTAAAATTG 7080

CTGTTAAAAC TTGTTGCTGT CTTTTTGTAC GGCCGTAATC TGCCTCATCA TCATCACGGA 7140

AACGAGCATA ATTGAGCAGG GTCGAGCCAT TCATCTGCTG TTTTCCGACT TTAATGGTTT 7200

GGGTTGGAGA CTCAGTCTCG GTAGCGTATA AATCATCTCC GACTGTAGCT TCTGTTAGGG 7260

GACGCCCATT CAATGTTGAA AATTGAGCAT CAATCGTCAC CCCATCAGGG AAAAGCGTGT 7320

CAATCGCTGT GGCAAAGGCC TGGAAATCAA CCAAGGCGTA GTACTTAATG TCCAAGTCAA 7380

AATTATCTTT CAAGACTTGG CGAACCATTT CTGCCCCTTT TTGCCCCTCT TGTTCTCCTA 7440

ACTCGTAGGC TACGTTTAAC TTGTTATCTG TCTGTTTTCT ACCATTAATC ACTTGACTAT 7500

AACCATCTAT ATAGACCAAA TTATCACGCA TGAAACTGAC TAGCTTCATT TTCTTATCTG 7560

AGCCCCCGAC ATTTAATACC ATAATAGAGT CAGTTCGTGT CTCAACACTG TTCTGGCCGA 7620

TTCGACCATC AGTACCCATG ATTAAAATAT TAACTCCATC TCTAGTGTCC TGACCATTAA 7680

AGACTTCTAC TTGAGCTGCC CGGGCATCAG CAGTTTTCTT TGCGCTAGCA TCTTGGTAAC 7740

CACGCAAAAA CATGAATACC ATGGCCAAAG CCACACAGAC CAAAAGTGAA AAAATCACCA 7800

TAAAAATTCG TTTAAGACGG AGCTTCCGTC TTTTCTTTTT TGGAGGGAAA GAGAGTGCTT 7860

GTGATTTGGA TTGTGAGCGA CTCCGGTTCG CATAGCTTGG TAAGTCAACC TGCTCTTCTC 7920

TTTCTTGTTC CAAGCTAGAG CTACTATTTC CCCTAGCAAG AGTTAGCTTT TCTTGCAAAT 7980

AGGCAAACTC ATTTTTTTCT CTCTCATTGA GATAGTGAAT ATTTTTTAGC AAATAATCAT 8040

AACGCAACTG CTCATGATGA CTTAAGGGAT TTTCTTTACT CATCTTCTCT CCTTTCCATG 8100

GTCTGATATT GGATAAATAG GATAGGCACC CAGAATTTTA TACTGGATTC CAATCGCTTC 8160

TAATTCTTTT TGGGCAAAGT GGACCAAGTC CTTATCGGTA TAATCCACAT CGATAATGAA 8220

AAAGTATTCA CCCAGTGCTG TCTTGAGTGG ACAACTTTCA ATTTTTGTCA AGTCAATTCC 8280

TCGCCAAGCA AAGGTCGACA GGGCCTTATA AAGTGCACCT GGAAGGTTGT CAGGTAATGT 8340

CAAGGCCAAA CTCATCTTTT CAGTTTGTGC TTGCAAGGGA ATACTAGGCT TTTCAGCTCC 8400

TAGAACCCAG AAACGTGTGA AATTGGCTTC CATTTCCTGA ATATCCTCGG CAATCAGTTC 8460

CAATCCATAT TCTTCAGCAG AACTTCTAGG TGCAACTGCT GCAAAGGGCT GGTCTGGATG 8520

TTCGGAAATA AAACGGGCCG CATAAGCTGT ACTAGCTGTT ACCTCGATTT GAGCCTCTGG 8580

ATATTGTTCA TCGATGAATT TCTTTCCTTG AGCCAAGGCC TGTGGATGTG AAAAAATCTT 8640

TTCAATCTTA GTATGGCCTG GAACCACCAT CAACTGCTGA TGAATAGGCT GAACGATTTC 8700

TGCTACTGCT TGGATGTGAG CCTGATGAAA AAGATAGTCC AAGGTTTCAT GAACACTACC 8760

CTCAATAGAA TTTTCAACTG GCACCACAGA ATAGTTCACT AATCCTTGCT CATAAGCCTT 8820

GATGACATCT GTAATGTTGG CAAAAGCCTG CAATTCCTCA TGAGGAAAAG CTGTCTGCAC 8880

AACGTGGTGT GAAAATGATC CCTTGGGACC TAGATAAGCA ATTTTCATCT TAGTTCCTCT 8940

ATAATTTCCT CTGGGCTTAG CTTGGTCACA TCCAAAACCC GACTAGCCAC TTCCTCATAC 9000

CAAGCCTGTC TTTCTTGGAA AATAGCTACT AGTTCTTCCT TGCTATTATT TAGAAAAAGC 9060

GGTCGCTGAT TGTCCTTATC AGCTGCGATA CGTTGGTAGA GGGTTTCAAA ATCTGCTCTC 9120

AGGTAGATGT TATCTGTATT AGTCTTGAGT AAGTCACGAT TTCTCTGAGA AATAACCACT 9180

CCTCCTCCAG TTGACACGAC TTGGTCTGTT TGTAGTAAAT CAGCTAGGAC TTCTGATTCT 9240

ACCTGACGAA AGGCTGTTTC TCCCTTTTCA GCGAAAAAAT TCGCAATGGA CATACCTAGG 9300

CGATTCTCAA TCAGAGCATC CATATCAAGG TAATTAGGGT CCAAGCCTCT TGCAATAGTC 9360

GATTTTCCAG CCCCCATAAA CCCTAATAAC ACCTTAGCCA TGAATCAAGC TCTCCAAATC 9420

ATCAAAGAAA CTAGGATAGC TGGTATTGAT GGCTTCTGCA CGGTCAAGCT CCACCTCTCC 9480

ATCTGCAACC AAGAGGGCTG CGATAGCTGT CATCATGCCG ATACGGTGGT CACCAAACGT 9540

ATTGACTCTA GCACCGTGAA GAGCTGATTT TCCTTTGATA ATCATCCCAT CTGCCGTAGG 9600

AGTAATATCT GCTCCCATAC TATTTAAGGC GTCTGCCACA ACCTGAATAC GGTCTGTTTC 9660

CTTGACCTTG AGCTCCTCAG CATCCTTGAT AACTGTTACA CCTTGGGCTT GGGTCGCAAG 9720

CAGGGCAATA ATGGGCAATT CATCAATCAA TCGTGGAATC AAAGCGCCAC CAATCTCTGT 9780

TCCTTTCAAG TCAGAAGACT CAACAATCAA GGTAGCAGAT TTAGCGACTG GATCGATTTC 9840

AGTTATTTCC AATTTTCCAC CCATGGCACG AATGACATCA ATAATACCGG TGCGAGTTTC 9900

GTTGATCCCC ACATTCTGCA GCACTAGACG AGAATTTGGA GCAATCAAAC CTGCGACTAA 9960

CCAAAAGGCT GCACTGGAAA TATCTCCTGG TACGACCACC TTCTGTCCTG TCAATTTTTG 10020

TGGCCCCTGG ACTGTGATTT TCTTACCATC CACACTTAAA TGACCACCAA ATTGTTTCAA 10080

CATATCTTCA GTATGATTAC GGGTGTACTC TTTTTCGATA ATAACTGACT CCCCCTTAGC 10140

TTGTAAGGCT GCAAACATCA AGGCTGACTT GACTTGGGCA GAGGCAATTG GCAACTCATA 10200

ATGAATAGGT CTTAGGTTTT TCGTCCCTTT TAAGCGAAGG GGAGGCAAGT CTCGTTCAGT 10260

TTGCCCTGAA ATGCTGACGC CCATTTTTTT CAGTGGAAGG GTCACACGGT CCATAGGACG 10320

TTTGGAAAGA CTATCATCTC CAAACATCTC TACTTCGAAA TCTGCACCAG CAAGGACACC 10380

TGAAATCAGG CGAATCGAGG TGCCAGAATT TCCCATATTA AGGGCATTTT GTGGCGCTTT 10440

TAAGCCAGCC ATGCCTACAC CTTGAATGGT AATAACCCCA TCTTTATCCT CAATTTCAAC 10500

ACCAAGGTCA CGAAAAACCT GCATGGTCGA AAGAACGTCT TCACCTCGCA GAATATCATA 10560

AACCTTGGTC TCACCCTCAG CCAAACTTCC AAAGATAATG GAACGGTGGC TGATAGACTT 10620

GTCACCTGGG ACGCGGATAC TACCATGTAA ATGGCGAATG TTTGTTTTTA GTTTCATACT 10680

GGACCTCATA CTTGCAATAC TTTTACCTAT TTTATCATAA AAAGCCAGAA ATTCCTTAAA 10740

AATTCCTGAC TTTAGGATCG TTCTTTTCTT ATTTCAGCAA TTCTGAAACT GGTTCAAAAA 10800

CAATTTTTTC AATATCAGAA AGGTAAATGG CCAATTGTTG TTGCTTGGTA AAGAATTCTG 10860

ACAAGAGGCT ATTTCCTTGA ATCTGTTTAC CAAAGCCTTC CATCTTAGCT TGGAAGGACG 10920

CATCTGGCAT TTGACCTGTC TGTGCTAGTT TTTGAATTTC CTCTTGAAAG GCAAGATAAT 10980

CTGTAAAGAT TTTGCTTGCC TCAGCATCTG CTGCAATCGC ATCTTTAGCT GCTTTAACAG 11040

CCTTGTATTC TGGTAATCCG CGTAGACCGC GACTGAGTTC GTTTGCACTA TCGTAAATAT 11100

TTGACATGTT CTTCTCCTTA TTTGATGACG ACTGTATAGT CAGTATTTTC TGTTATGAGA 11160

TGCTCAGCTC TTTCCAAGTC TTGAGCATTT TTAAATGAAA TTTGTAGGAT TCCGTGAATA 11220

TCCTCACGAT TTTCCTCGTT GATGTGGATA TTAACCAAGG AAGTTCCACG TAGCAGTTCC 11280

AAAATCCGCA GGATGACATC TTCTTCATCA GGAACGTCAA CATAGAGGTC GTAAGAGCTA 11340

TCCACACCAC CACGCTTATG GATTTCCATG GTCTGGCGTT GTTCACGCGC TTGGTTAAAA 11400

AAGTTCCAAA TTTGCTCTTC ATCTCCCTTA CTAATGGCCT GACCAATCGC TTCCAAACGT 11460

TCCTTGAAAT CCTCAATTCT ATCCAGAATG ATCTCGCTAT TGGACAAGAG AATGGAGGTC 11520

CACATTCCTG GCTCGCTTTC CGCAATTCGG GTCATATCTC GAAAACCACC TGCCGCAAAG 11580

CGCCTTGCCA TCTCATGCTC TTGAGCATAG ACCGCAGTCT GCTCCATGAG ACTAGAAGCC 11640

AAAATATGAG GAAAATGGCT AATCTGAGAA GTGACACGAT CATGCTCCTT GGCATCAATC 11700

TCGATAAAAC GAGCATGAAG ACCTGAAAGC AGATCCTTCA TTTCCTTAAG CGTGTCCTGA 11760

CTTGTCAGGC TTGAAGGTGT AAAGATATAA TAGGCATTTT CAAAAAGATT GACATCTGCC 11820

GAAGCAGCCC CTGTCTTGTG ACTACCAGCC ATGGGATGGG CCCCGACAAA GCGAACAGAC 11880

TTGCCAGCCA AATACTGCTC CGCCGCATCC ACAATGGTTG ACTTGGTCGA ACCAGCATCT 11940

GAAATAATAA CGCCTTCTCG CAAATCCAAA TTGGCCAACT CCTTAATGAA AGCAATAGTT 12000

TGTTTGATTG GCAAGCTGAG GATAATGACA TCTGCCAAAG GAGCAAAACT AGCAAAATCA 12060

TCCGTTGCAC GGTCAATCAT ACCTTCTTTC AAGGCGATAT CTCTCGAAGC TTGACTACGA 12120

TTATAACCTA AAATTTCATA ATCTGGATGA TCGCGTTTGA TACCAAGTGC CATAGAGGCT 12180

CCAATCAACC CAAGACCTGC GATATAGATT GTTTTTGCCA TAGGAACTCC TTAATAGTTC 12240

TTTGTATAGT CTCGGTGTTT GGCTACCGCT TCTTTTAGTT CCTCAAGATT ATCTGATGAG 12300

AATTTTTCGA GGATTTCTTG CGCCAGAACC GTTGCTACAA CTGCTTCCAT GACCATTCCT 12360

GCAGCTGGAA GAGCAGTCGG ATCACTTCTC TCCACGGTTG CCTTGTAAGG TTCGTGGGTT 12420

TCGATATCCA CACTCATAAG AGGTTTATAA AGAGTAGGAA TGGGTTTCAT GACCCCACGA 12480

ACAACGATGG GTTGCCCATT AGTCATACCA CCTTCAAAAC CACCTAGATT ATTGGTACGG 12540

CGAGTATAAC CGTCTTCTTT AGACCAGAGA ATTTCATCCA TAACTTGGCT GCCTTTACGA 12600

TAACCAGCCT CAAAGCCAAG ACCAAATTCC ACCCCTTTAA AGGCATTGAT AGAGACAACA 12660

GCTTGAGCCA ATCTTGCATC CAATTTTCTA TCCCATTGGA CATAGGAACC AAGACCAACT 12720

GGAACGCCTC CGACGACTGT CTCCACAACC CCACCGATGG TATCACCATC ACGTTTGATT 12780

TGGTCAATAT AGTCCTTGAT TTCCTGTTCT CGTTCTTGGT TGACAATAGA AACTTCAGAC 12840

TGGGCAGCTC TTTGCTTAAT TTCAGCGACT GTCAGATTTT CAGGAACATC GATTTCCTTG 12900

CCACCAAAGA CCACGACATG GTTGGCAATC TCCATATCCA GCTCAGCCAA GAGGCGTTTG 12960

GCTACTGCAC CAACTGCCAC CCGCATGGTG GTTTCACGAG CTGATGAACG CTCCAAAGAA 13020

TTTCGCAAAT CATCAAAACG GTACTTAATC CCCCCAACCA AATCGGCATG ACCTGGGCGA 13080

GGATGAGTAA TTTTCCGCTT GCTTTTAAGG CGGTCTTCAA TGTCCTCCGC AGACATGATG 13140

TCCAGCCATT TCTGGTGGTC CTTATTGATG ACATCCATAG TAATAGGCGC CCCTGTCGTC 13200

TTCCCGTGGC GAACGCCCGA AGTAAAGACA ACCTGGTCAT TCTCAATCTT CATACGACCA 13260

CCACGACCGT AGCCACCCTG ACGGCGTCTA AGGTCCTCAT TGATATCCTC AGCTGTCAAT 13320

GGAAGTCCAG CTGGAATTCC CTCAATAATA GCTGTTAGAC GGGGGCCGTG TGATTCTCCT 13380

GCAGTTAAAT ATCTCATACA CTCTCCTTAT TTTACCAAGT AGTCTTTCAT CTCTTCCAGA 13440

GAAACTGGGT GAATGGTCGC TGAACCAAGC TCTGGCACCA AGACCAATTT CAAGGTGTTA 13500

CCACGCGCTT TCTTGTCATG AGTAAGAGCC TGATAAAGCT TGCCAACTTC CCAATTTTCA 13560

TAGTCAACAG GCAAACCGAA TTTCTGACAC ATCTCTGTGA TAGATTGGGT AATGCCAGCT 13620

GGCATGAGGC CTTTTTCCTC AGCAACCTTG GAAATCTGTA CCATTCCCAT GGCAACAGCC 13680

TCTCCATGCA TGACCTTGCC ATAACCGGCA GTCGCTTCGA TGGCATGGCC AATAGTGTGG 13740

CCAAAATTGA GGTAAAGACG AATACCATTG TCCAACTCAT CTTCAACCAC CATCTTGCGC 13800

TTCACCTGAC AAGAATGTTC AATCAAGGTC TCTGCATGTT CCAAAATACT CTCAACAGAA 13860

CCATTCAGTC CCGTCAAGAG AGCCCACAGT TCTGGATCCT CAATCAAGCC ATACTTGATA 13920

ACTTCACCCA TCCCTTCAAT CAACTCTCTT TTTCCGAGGG TTTCAAGAAC AAGTGGATCA 13980

ATCAGAACCC CATCTGGTTG GGCAAAGGTC CCCACCATAT TTTTAGCAAA TGGTGTATTA 14040

ACGCCTGTCT TTCCACCGAT AGAAGAATCA ACCTGAGCTG TCAAACTAGT CGGAATCTGA 14100

ACAAAGTGAA TACCCCGCAT ATAGGTAGAG GCTACAAATC CAGCCAGGTC CCCAACAACG 14160

CCACCACCAA GAGCAACGAT TCCATCGCTA CGAGTCAGAC CTTGCTTGAC TAGAAATTCA 14220

TAGACTTTCT GAACAGTAGT TAAATTCTTT CTTTCTTCAC CTTCTAAGAA ATCAAAAACA 14280

GCTACCTGAA AACCAGCATC TTCTAGGCTG AGCTTGACCT TCTCTGCATA GAGAGAGGCT 14340

ACATGGTTAT CTGTCACAAT GACTACCTTT TGCGGTTGCC AGAGTTCTCG CAACCACTGA 14400

CCAGCCTGGG CCATACAACC TTTTTCAATC TGAATATCAT AAGGATGGTG AGGAATATCG 14460

ATTCTGATTT TCATAGGAGA GTCTCCCTTT CTTTATTGGT ATTTTTCTGT TAAAGACTGC 14520

CAAATCTCTT CTGTCGGCAT TTCCTTGCCT GTCCACAGTT GAAAAGCTTC TGCAGCTTGA 14580

TAGAGTAACA TTCCCAGACC ATTGACTGCT GGATTGCCCT GACTTCTAGC CCATTTCAAA 14640

AACGGTGTTT CAAAGGGTTG GTATATGATA TCTGCAACTA AAAGAGTTTC TGGTAAGACT 14700

ATGTTTTCAG GAACAGGAGA GGATTGGCCA TCCATGCCCA CACTGGTGGC ATTAACTAGC 14760

AAATCCGACT CGGCAATCCT TGCTTGCAGT TCAGAAACAT ATTCTAAAGC ACACAAATCC 14820

ACTTTAAAAC CTGTCTGCTC CTGTAACTTG TCTAGGTAAG GTCTTGTTTT TTCCATAGAA 14880

ACGGAACGAA CAAAGACCGA AATCTGACTG ACGCCATCCA AAATAGCCTG TGCCAAGATT 14940

GATTTAGCCG CACCACCTGC ACCCAGCAGG GTCATCTTTT TACCTGAAAT TGTAAAAGAA 15000

GGCAAGCACT TAAAAAATCC CTTGCCATCT GTATTATATC CAATTAAATT GCCATTCTCA 15060

TTGACAACCG TATTAACCGC ACCAATCAAG CGCGCTTCAT CGCTCAGCTT ATCCAAATAA 15120

GGAATCACCT GCTCCTTATA GGGCATGGAC AGATTGATGC CAAACATCTG GTAGCGACGA 15180

ATATTGGCCA CTGTTTCTAC CAAGTCACTC GCTTCAATCT CCCAAGCCAC ATAAGCACCG 15240

TTGGTAGCTG TCGCCTCAAA GGCTCTATTG TGGATGAAGG GAGAAATAGA ATGCTTAATA 15300

GGATTGGCAA CAACTGCAGC TAAACGTGTA TAGCCATCAA GCTTCATCCA AAATCTCCCT 15360

GATTTTTTTC ATGCTAGCTA GAGAAATCTG CCCAGGGGCA CTAACCTCAT CCAGACTGGC 15420

AAAAGACCAA CTCGAACCAG TCACATCCGC AGTGATACGA GAGACCTTGC CCACCTTACC 15480

CATAGAAATG GTCACATATT CCTGTTCAGG ATTGAGGGTT TTAAAGCCTC GTGTATAGTT 15540

CATCAAGTCT AAGACATCCT GCTCCGTGTG AGCCATCACC GCAACCTTAA CAAGTTTTGG 15600

ATTTAGGATC GTCAACTCTG ACAAGATTTC CATCATGTTC TCAGGTGTTT CTTGGAAATT 15660

ATGGTAACTC AAAACAAGAT TTGGGAAGTC CAGCATTTCC TCAAAAACAT CCTTGTAGCT 15720

ATAGTACTCA AAATCAATAT AGTCTGGTTG ATAGAGTTGC GCAACTTCCT TGATTAGATG 15780

GATATACTCT TCTGGAGAAA GGTCGATTTC TCCACCTTCG GAGCGAGTTC GTAGCGTGAA 15840

AACCAACTCA CGGCCTGCGA ATTTTTCAAA AATGGCTGGA GCTACCTGCA AAATCGCTTC 15900

TTTAGGCAGA TAGTCGGCAC GCCATTCAAT GATGTCGGCA TCCAGGTACC TCGTGGCATC 15960

CAGAGCCTGA GCCTCCTCTA AACTTCTTGG CATTACTGAA ACGATTAATT TCATTTACTA 16020

ACCTTCATAC TAATCACCTT GAGGTAATTA CTACTTTCAT CTTTTTTATT ATAGGCAAAA 16080

TCTGCTGGAA GACCATATTT GTTTAAAATC TGGTAACTTC TTCCTGCAAA ACCTTTATCA 16140

ATTTGTTCTG TAAATTTCTG ACGGGAAACA TTGGCAGCAT TGGTACTGGC AATGATAATC 16200

CCTCCCGGAT TTAAAATCTC AAGACTCTGG GAAATCAACT TGTGATAATC CTTGGCCACA 16260

GAGAAAGTTT GTTTTTTATT CCGAGCAAAG CTAGGCGGAT CTAGGACAAT CACATCGTAG 16320

GTCAAGTCTT TGCGTTTGGC ATATTTGAAA TACTCAAAGA CATCCATGAC TATAAAACGA 16380

TGCTCGTCTG TGCTGAGCCC ATTTGCCTGA AAATGCGCTT GAGACAATTC TCGTGAACGT 16440

TTGGCTAGAT CAACAGAAGT TGTATGGCTA GCTCCTCCCA TGGCCGCAGC TACTGAAAAA 16500

GCCGCTGTGT AGGAAAACAT ATTGAGTAAG GATTTACCCA TAGCCAAGCC GTCAACTAAA 16560

CTACCGCGAA CCTCATGCTG GTCTAGGAAA ATTCCTGTCA TCAAGCCATC ATTCATAAAG 16620

ACTTGATACA GGACACCATT TTCTAAAACA TTGAAAAAGT CAGGTGCTTC TTGACCATAA 16680

ACATGGGCAG ATTCATAGTC CAAACCCTTA AAGCGGATTT TCTCATAAGC TCCTAAAACC 16740

TCAGGGAAAA CCTGTCTAAA GGCTTCTGAT ATAGTCTGAC GAATCTGATA AACATAAGAG 16800

TTATACCAAG AAAAGACGGC GTAGTCGCCA TAAAGGTCCA CTGTCAGACC CCCAAAGCCA 16860

TCTCCCTCTT GATTAAAGAG ACGAAAGGCA GTTGTCAAAT CATCTTGATA GTAGGCGTTT 16920

CTCTTTTCTT TGGCTTTTCT AAACAACGTT TCAAAGAAAG CTTGATTGAA GGCCACCTTG 16980

TCTTTGCTGA TAAACCAGCC CAAGCCCTTG TTTTGCTGAG AAAGGTAGGC AGTCCCAAGA 17040

AAGTTTCCTT CCTGACCCTG CACCTCTACT TCCTGATCCT TAAGATTGAC ATTCTCAAGA 17100

TCACTGGCTT CTAGTAAAAC TAGCCCCTTA GCAAGCTTCT TTTCAACCCT TTTGCTGACT 17160

CTTATTCTAT TCATAACTAC CATTATATCA AACTTTTAGA CAATTCTCAA AAAAGAAACT 17220

ACCCTTGCTT TTTTACTCTT CTTTTAAAAA ATGGTATACT AGACTTCCTG CAAAACTAGG 17280

AAGTAAATGT GTAAGAATCA CAGTAAAAAA TGCTCTTCCG TCTTGGAGGA GCATTTCTTT 17340

TTATCAACGA AAATCAAATA GCAAACTATG AAACTAGCCT CAGGTTAACT GTGAGATTAT 17400

AGGTAGAGAG GTTGTATCAG CAATATGTGT CTGTCAAATT TAGTGACAAA GGTAGTAGAA 17460

GAAAGATAAA GAAATAAATC AGCTTCAGTA GGTATCTGGA AAATTTGATT TTATAGAGAA 17520

GCCTTTTGTT ACAAACTCAA TATACTATCA ATAAATAATA TTATAGAAGC AACAATAATT 17580

ATAATTTCAC CTATCTGCAT CATTCTATTT CGAACTCTAA ATATATGTTC TATCAAAAAT 17640

ACTTGGAACA CACACATTAT AGGAATTAAC GTTTTTGAAA TTGAAAAATA TCCAAATAAA 17700

TAAACTATAA ACAACAAAAA TAGAACTATG TTATATTTCT TATTCAAAAC ATTCCTCCCT 17760

ATATATTTTT GATTACCAAT CTTAATCATT TACAACTACA TTCTAACAAA CTATAAAAGC 17820

GTTTGTCGAA TTGAATTTAT CAAGCAAGCG ACCAACCAGT TCATCTTTTT TCTATTTCTG 17880

CCAATATGCG TGACAGGTAA TAATGATAGC CAAAAATAGC AAGAGCAAGC AAGACGATAA 17940

GAGCTCCTAC TCCCAAGCTG ATGGCAAGGA TAGGGGAGAG AGACTGAACC AAGAATATGC 18000

TCCCAATTAC AAGGGCCATC AGGATTGCAC TATAAATAAA CAATAAAACT ATGGCGACTA 18060

TGCCATTTGA ACGATTCACC AGGTCCGTAA TGCTACTCCA ATTGGTTGAC AGATTTTTAA 18120

CGTCCTTAAA GTAATGGTGG CAAGAAAGGA TGACACTGGC AATGATCCAG ACTACAAGAA 18180

GGTAAATCAT CGAAATGATG GGCAAGCCTA GATATAGAGA AAGACCAAGC AAAGTCAGAA 18240

CTGGTAAAAA GGACTGGACA GCATATATAA TCCAAAATTT CACTTTCACA TAACGAGCAA 18300

AGTCAAAGGG TAAACTCTTA AGAAAATCAA CATTTTCCCT CTCCAAGGAC AAGGCAATTG 18360

AATGCAGGCT GGTGATATTG TTATTGACAA CTGCTATAAA GAGAGCTATA AAAAACAAGG 18420

GTAACCAGTA TGGAGGATGA ATGTCTGGAA CTATCTGAGA ATCTCGGATT TTGGAAATCA 18480

GACCGATCAT CATGAGATAA GGAAGGAAAG CACTTGTAAA AAGCACTGTA ATCACGCCAG 18540

TCCCCTGTCC CAAGAGGGTG AGGTGGTAGC GTAAAACCAT GCGAAAAAAT CCCTTTTTAG 18600

TGGTTGAAAT TCTCTCCTTG CTGCGACGTT CTTTTTTGAC CTTCTCCTCA CTATTAAGCA 18660

GGATCACGTC ATAAAAACGA GGAAGGACCT TCTTTTTGGT CAGATAAAGC AGGAAGAGAG 18720

TTAGTCCTAT CCAAGCGAGC AGACCCACTA AGGCTTCTGT CGAAAAAGGC TCCACTGCTA 18780

TTTTGTAAAA GATATGAAGA GGATAAAGGA GAAATGGAAT GTCTCTAACT TTGTCAACAA 18840

TACTTCCAAA AGTCGACTGA AGAAAGAAGA TAAATATTAA AGGTATGAGA ACTCCTATCC 18900

CAATCATCAC ATTCGAAAAA ATAGACTGAT ACTTTCTGAA GACCCTAGTT TGAGCCAAGA 18960

AATGCACTGC CACTACCATC ACTAGAGCCA CAGAGACAAA TAATAAGGTC AAGGACAGTA 19020

GCATCAAAGG CAAACCCAGC CATAGAGAAG GAGCTAGCCT AATGTAGAGG ACCAGAAAAT 19080

AAGCTAGGAT TGGTACAATT CCAGTTAGAG CTGGCAAAAG GACAGACAGT CCTTTAGCAA 19140

TTATAATCTC TGATTCTTTA AAGGCATAGG GCCTATACGA TACCAAATCC TTACTCTCAT 19200

AAAAGACATT GTAAAAGGCC GTTAAAGAAG TTGAAAAGGC AATCACTAGT AAAATAGCAA 19260

TCATCGAGCT AAAATAAATA GGTATTTCCT CAAAAGGAAA ATGAATGGCT ATATTACTAA 19320

AACAGATGAT CATCAAGAGA CTGGAAAAAA TGTAAGAACT TAAGACTCTA GCGGAAACAT 19380

TTACTTTTTT 19390

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18436 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

CCGAGCGTCG TTACAGACTT TATCAAGATT GGACGCAAGA AGAAATTCAA CATATAAAGG 60

AAAATATGGC ACAATCTCCA TGGCATACTC ATTACCATGT TGAGCCAAAA ACAGGACTTC 120

TCAACGACCC AAATGGCTTT TCTTACTTTG ATGGCAAGTG GATCCTCTTT TACCAGAATT 180

TTCCTTTTGG TGCAGCCCAC GGTTTAAAAT CTTGGGCACA GCTAGAAAGT GATGATTTGA 240

TTCACTTTAA AGAAACTGGA ATCAAAGTTT TACCAGATAC TCCATTAGAT AGCCACGGTG 300

CCTACTCTGG TTCTGCCATG CAATTTGGCG ATAACTTATT CCTATTTTAT ACAGGAAATG 360

TTCGCGATAA AAACTGGATC CGTCACCCAT ACCAGATCGG TGCTTTGATG GACAAGGAGG 420

GTAAGATTAC AAAGATTGAC AAGATCTTGA TTGACCAGCC AGCAGACTCT ACTGACCACT 480

TCCGCGATCC ACAAATTTTT AACTTTCAGG GTCAATATTA TGCCATTGTC GGCGGACAAG 540

ACTTGGAGAA AAAAGGTTTC GTTCGTCTCT ACAAGGCTGT CAATAACGAC TACACAAACT 600

GGCAAGCAGT TGGCGACCTT GACTTTGCTA ACGACCGTAC TGCCTACATG ATGGAATGTC 660

CTAATTTGGT CTTTGTAGAG GAACAACCTG TCCTTCTCTA CTGTCCACAA GGATTGGATA 720

AGAAAGTTCT AGACTACGAT AATATCTTTC CAAATATGTA TAAGATCGGG GCTTCCTTTG 780

ACCCTAAAAA TGCCAAAATG GTAGATGTGT CTCAACTTCA AAACATGGAT TACGGTTTCG 840

AAGCCTATGC AACTCAAGCC TTCAACGCTC CTGATGGGCG TGCTCTAGCA GTTAGCTGGC 900

TTGGTTTGCC AGATGTTTCT TACCCATCTG ACCGTTTTGA CCACCAAGGA ACCTTCTCTT 960

TGGTCAAGGA ACTCACTATC AAAGACGACA AGCTCTACCA GTATCCAGTC GCTGCTATTA 1020

AGGACCTTCG TGCTTCTGAA GAAGCCTTCT CAAACCGTTC CCAAACCAAG AACACTTACG 1080

AACTTGAACT CAACTTGGAA GCTAATAGCC AGAGCGAGAT TGTCTTACTT GCTGATAAAG 1140

AAGGTAAGGG ACTTTCAATC AACTTTGACC TTGTAAACGG TCAAGTAACA GTGGATCGTA 1200

GCCAGGCTGG AGAACAGTAT GCCCAAGAAT TTGGGACAAC TCGTTCTTGC CCTATCGAGA 1260

ATCAGGCTAC TACTGCTACA ATCTTCATCG ATAACTCTGT CTTTGAAATT TTCATCAATA 1320

AAGGAGAAAA AGTATTTTCT GGTCGTGTCT TCCCACATGC GGACCAAAAT GGTATCCTGA 1380

TTAAATCTGG AAACCCAACT GGAACTTACT ATGAATTAGA TTATGGTCGC AAAACTAACT 1440

GATGTCGCCA AACTTGCAGG CGTCAGTCCT ACTACCGTTT CTCGGGTTAT CAATAAAAAA 1500

GGGTATCTAT CTGAGAAAAC CATCCAAAAA GTCAATGAAG CCATGCGAGA ATTGGGCTAT 1560

AAACCCAACA ACCTGGCTCG TAGTCTGCAA GGAAAATCAG CTAAGTTAAT CGGCTTGATT 1620

TTCCCCAATA TTTCCAATGT TTTCTATGCA GAATTGATTG ATAAATTGGA ACACCAACTC 1680

TTCAAAAATG GTTACAAGAC CATCATCTGC AACAGTGAAC ATGATTCTGA GAAGGAACGC 1740

GAATACATCG AAATGTTGGA AGCCAATCAG GTGGACGGCA TCATTTCTGG TAGTCACAAC 1800

CTAGGAATCG AAGACTACAA TCGTGTGACA GCGCCGATTA TTTCCTTTGA CCGAAACCTA 1860

TCGCCAGACA TCCCTGTCGT CTCCTCTGAC AACTATGCTG GTGGGGTTCT TGCTGCCCAA 1920

ACCTTGGTCA AGACAGGTGC CCAGTCTATC ATCATGATTA CAGGGAATGA CAATTCTAAT 1980

TCGCCAACCG GACTGCGCCA CGCTGGTTTT GCATCCGTAC TCCCAAAAGC TCCTATTATC 2040

AATGTTTCCA GTGACTTTTC TCCCGTCAGA AAAGAAATGG AAATCAAGAA TATCTTGACC 2100

CGGGAAAAAC CAGATGCCAT TTTTGCTTCG GATGATTTGA CAGCTATTCT GGTCATTAAA 2160

ATCGCTCAAG AATTGGGCAT TTCTGTCCCA AAAGAGCTCA AGGTCATCGG CTATGATGGG 2220

ACCTACTTTA TCGAAAATTA CTACCCTCAA TTGGCTACTA TCAAGCAACC TTTGGAAGAG 2280

ATTGCTTGTC TCACTATTGA TCTTCTCTTG CAAAAGATTG AAGGCAAGGA AGTCGCCACA 2340

ACTGGTTACT TCTTACCAGT TACGCTATTA CCAGGAAAAA GTATTTAAAC ACAAGAAAAC 2400

TCAGACCGAT TCGTCTGAGT TTTTATGATC TTAAATTTTC GAGATAGCGC TGGGCTGTCT 2460

CTAGGTTAAA GGTTTTATCT GAGATGAGGC GCTCTACTAG GGGAGCAACT TCAGATTCAC 2520

TAGCCCCAGC TAGGAGAGCT AGGGATTTGG CCTGTAGTTT CATGTGGCCT TGCTGGATGC 2580

CCGTACTTAC CAAGGCTTTG AGGGCTGCAA AATTTTGAGC AAGACCGATG GACACGATAA 2640

TCTGGGCTAA TTCTCTGGCA GAAGGATTTC CTAGTAGATC ATGACTGAGA ACTACACGTG 2700

GGTTGAGGCC GATAGAGCCA CCCTTAGTCG CTACAGGCAT GGGCAGGGTC ATCTCACCGA 2760

CCAATTCTTC TCTTTCAAGG TCCAGCGTCC AGCAGCTAAG ACCTTGATAG CGTCCATCTC 2820

GACTGGCAAA GGCATGGGCC CCAGCTTCGA TGGCACGCCA GTCATTACCA GTGGCAATCA 2880

AAATCGCATC AATACCATTA AAAATTCCTT TATTATGAGT AGCAGCTCGG TAAGGATCAG 2940

CCTGCGCAAA CTGACTAGCC AACGCAATTT TCTCCGCAAT CTCTCGTCCT TGATCCTTTT 3000

GGCGGCTCAA GTAGCGAAAG GCGATGCGAC AGCTTGCAGT CACCAGAGAA TCGGTCGCGT 3060

AGTTGGACAG GATTCCCATG AGACTCTGTC CCTGACTGAG TTCTTCTAAG ACTGGTTTCA 3120

AGGCTTCCAG CATGGTGTTG AGCATATTGG CACCCATGGC TTCCTGGGTA TCGACATGAA 3180

TATAAACAAC GAGAAAGTCT GGTTCGCCTT TTATCTGCTC GACATGCAGA TCACGCGCCC 3240

CACCTCCACG TTTAACGATA GAAGGATAGG CTTGATTGGC AAGCTCCAAG AGCTCCGCTT 3300

TCTTGCTGGC AATCTTCTCT TGCGCTAGTT TAGGATTAGC AACTTGATAA AGGGCTACCT 3360

GCCCAATCAT CTGTCGCTGA TGGACTTGTG CAGTAAAACC ACCTGCACGC TTGATGATTT 3420

TGCTGGCATA GCTGGCCGCC GCAACCACAG AGGGTTCTTC TGTCACATAG GGAACGGTGT 3480

ATTCCTGACC GTTGACAAGT ACCTCCGGAA CCAGTGAATA AGGCAGAGAA AAAGTTCCCA 3540

CTACATTCTC ACTCAGCTGG TCTGCCACAG TCACGCTCAT CTGTTCATCC TTCTCCAGAC 3600

TAGCTTGTCT CTCAGGACTA AGGAGCGCCT GAGCTTTTAA CAGCTCGAGG CGCTCTTGGT 3660

ATGATTTTTT AGAAAATCCA TTCCAACTTA TCTTCATTAT TTTTCAACCT TGCTATAACG 3720

GCGTTGGTGG TCGAGAATTT CAACCAAGGC AAAATCTTGA TTTTCATAGC CAGCAAACTG 3780

GGCAGAGTTA GTTTCATCCA AGTTTACTTC CTCAAAAAAG ACCTTTTCAT AGTCTGCAAC 3840

GGATAGGGCA GTTCGTTGGT TGAGCTTGTT CAAACGGTCT TTATCCAAAT AAGCTTCATA 3900

TCCTTCAACC AATTCACCAC TGAAGAACTC AGCCACAGCT CCACTTCCGT AACTATAAAG 3960

GGCGATTTTA TCCCCAGCTT TCAAGCTATC TGTATTTTCC AAGAGAGACA AAAGTCCAAG 4020

GAAAAGTGAA CCTGTGTAGA TATTCCCCAC CTTTTGACTG TAGAGAATAG ACTGGTCAAA 4080

ATGCTTTTGT AAGAGGTCTT TTTTCTCTTG AGGCAGGCTC TTATCCATGA TTTTTTTCAA 4140

GCCTTTTAGC GCTAATTTAG GATAAGGCAA GTGGAAACAA ACAGCCGCAA AATCATCCAA 4200

AGTAAGCTGG TAGCGTTTTT GATATTCAAG CCAAGTCGTT TTCAAACTAT CCAAGTATTG 4260

TTGGGTAGAA TAGACACCAT TTACATAAGG AGTTGTCGAG TAATTTGGTC GCCAGAAATC 4320

CATGATGTCA CGGGTCTGAG CTACATTGTC ATTATTAAAG GCCATCATGC GTGGATTTTG 4380

TGTAATCAAC ATAGCTACAC TTCCAGCACC TTGAGTTGGT TCTCCTGGAG TTTCAATACC 4440

GTATTTGGCA ATATCACTGG CAATGACCAA GACCTTGGAC TCCGGAGAAT TTTCCACATG 4500

CAATTTGGCA TAATGGAGGG CAGCAGTCGC TCCGTAGCAG GCTTCTTTAA TCTCGAAACT 4560

ACGAGCAAAG GGCTGGATGC CCAGCAAGCC ATGCACAAAG ACGGCCGCAG CCTTACTCTG 4620

GTCAATTCCT GACTCGGTCG CCACAATGAC CATGTCAACT TCTTGTCTTT CTTGCTCAGT 4680

TAAAATAGAG TCACTAGCAC TGGCCGCCAA GGTCACGATA TCCTCAGTTA GGGGCGCAAT 4740

ACTCAATTCC TTGAGTAAGA GTCCTTTACT TAATTTTTCA GGGTCAATTC CCCTCGCTTC 4800

TGCTAAGTCT TGTAATTTCA AGACATATTG ACTGGTCGCA AAACCAATCT TATCAATACC 4860

GATTGTCATA TTTACCTCTG TTTTATCATT CATGTAAAAA ATCGTTCTAT ACTATTTTAT 4920

CACAAATGGC AGTAAAAGAG AGAAAAAAGA CTTGATTCAC CAAATCAAGC CTCTTATTGG 4980

TCATCATTTT AAAGAATGAT TAGTTGCTAG AGAGTTCACC GATATAAGTA GCTTTATAAG 5040

CTCCATTCAC AGTTATCAGC TCCTGGAGGA TCAAATTTCC TGAGTAAGTC CTTCCCATCT 5100

CATCTACAAA TTTTTGATAA AACTGACTGG TCGGAATTTC TCTGACATCC TTATCAAATG 5160

TCTTATCAAG TGTTTTACTA ACCTTCTCAG CAATCAATTG ATGCTCTTGC CATCCACTTT 5220

GAAACTCTGA GCCCGAACTA GAAACCATGA CTGGGATAAA CAACAAGGTC AGTAGATTTA 5280

CAGACAATAA GGAAAGTAGT AGACTTCCTG CAAAACTAGA ATCCTAGTTC ATGATTGATA 5340

ATACCAGCAA TCAAATTCAT TCGTAATCCG AAGCGTTTAC GATGATTTCG ATAGGTTGTT 5400

GAAAACATTT TAAACGTTTT TACTTTGGCA AAGATGTTCT CAACCTTGCT TCTCTCCTTA 5460

GATAGCGCAT GGTTACAGGC TTTATCTTCA GCTGTTAGCG GCTTGAGTTT GCTGGATTTA 5520

CGTGGAGTTT GTGCTTGAGG ATATATCTTC ATGAGCCCTT GATAATCACT GTCAGCCAAG 5580

ATTTTACCAG CTTGTCCGAT ATTTCTGCAA CTCATTTTGA ACAACTTCAT ATCATGACTA 5640

TAGTTCACAG CGATATCCAA AGAAACAATT CTCCCTTGAC TTGTGACAAT CGCTTGAGCC 5700

TTCATAGCGT GAAATTTCTT TTTACCAGAA TCATTCGCTA ATTGTTTTTT AGGGCGATTG 5760

ATTTTTACTT CCGTCACATC AATCATTATC GTGTCCTCAA AGCTGAGAGG AGTTCTTGAA 5820

ATCGTAACAC CACTTTGAAC AAGAGTTACT TCAACCCATT GGCTCCGACG GATTAAGTTG 5880

CTTTCGTGGA TACCAAAATC AGCCGCAATT TCTTCATAAG TGCGGTATTC TCGCACATAT 5940

AGAAAGCGTT ATCAATTTAT TTATCTCATT TTTCAGAAAA TTCTTTTATT TCTGTAAAGT 6000

CTACGATACT CGATGTGTTT TTATATAATG ATAGAGTCTG AGAATCACTG TTCCGCTAGC 6060

CATTCCAATA GAGATTACCA AAGCCAACAT GACAACCAAG GTCGCACTTG CCAGTGCTTT 6120

ATTATAGTCC CCTGTCACAA AAAAGGCAGT TGTTCGGTAG GAGAGATAAC CTGGAACCAG 6180

CGGTGCCAAA ATGGCCAAGA TAAAGACCAC AGCAGGTGTC TTATAAAGAA TACTTAAAAT 6240

CTGGCTGACA CAAGAACCAA TAATGGCTGC AATGAAGGTA GCTACAATGA CATTGGTCGG 6300

TTCCTTGAGC AAGAGATAGA TTAGCCAGAC AGTCATGCCC AAAATCCCTC CAGGTAAGAG 6360

CATAGACCGT TGCACATTGA GTACGATTAA AAAAGTGATA ATGGCAAGAA AACTTGCTAC 6420

TGCTTGTAAT AAAAAGGTTG TTAGTGTCAT ATTAGTTCAT CAATACCAAG GCGACAGAAG 6480

TTCCTGCCCC TAAAGCGAGG GTAATGAGCA GGGATTCAAA CATCTTACTC ATACCAGAGT 6540

TTATGTGGTT GGTCATAATA TCACGGACCG CATTGGTCAA GGCAATACCT GGTACAAACG 6600

GCATGACCGC ACCAGCTATA ATCAAATCTG CCGTTGAAGG AAAACCTGTG TAGCGAGCCC 6660

AAAACTGGGC AATTATCCCA AAGACAAAGG CTCCAGCAAA GGCTGTCACA AAGGGAATTC 6720

GGATAAATTT TTCCACATAG AGGGAAAAGG CAAAACCAAA TAAGGTCGCC ACTCCTGCCC 6780

CAAGTGCGTC GTAGATATTT CCGCTAAACA TAACTGAAAA GAAAGGAGCA CTAAAGGTCG 6840

CAGCCAGAGT TACCTGCAAC TTAGTATAGG GAAGGGGTTT GGCTTGCAAG GCCGTCAATT 6900

GCTTAAAGGC TGTTTCTAAG TCAATCTGCC CCCCAACTAG CTGACGAGAA ATCTGGTTCA 6960

CATCGCAGAC TTTTTCGATG TTATAAGAAG AGGAGGTCAC GCGCTTCATG CGCAAATATT 7020

GGTATTTTCA ATAGAGAAAA AGATAGCGGC AGGCATGGCA AGGACATTGC AATCCACAAT 7080

CCCCTGCGAA TGCGCGATTC GAATCATGGT ATCTTCTACA CGATGGATTT CTGAGCCACT 7140

TTTAAGGAGA ATAGTCCCCG CTAGCATAAT CACATCAATG ACGGCATTTA ATTCTTTTGA 7200

TTCTTCCATG CTTTCCTCCT TTTATCAACT CCCTCTATTC TATCACAAAT CCGGACTCAA 7260

AAAAAATCTT TGCCATGAAA TCATGACAAA GATTGATTAC TCATTTTGAT TATCCATCTG 7320

CTTTTAAGGA GTAGCTGAAG TTGTTTTAGG TTTGTAGATT GAAATCTTGA CTCTAGTCTT 7380

ATTGAGGTCT ACCTTTTCAC CTGCTCTAGG ACTTTGTTCA ACAACCATGC CTTCTGCACT 7440

ACCTGCAGGC GCTGTCGTCA CTTCTACAAC TTCTATATTA GCTTCCTTAA TCCCAACAAT 7500

TTGAATCAAA TTGTTCTTAG TAAACTCCAA GCTAGAACCA ATGTAACTCG GCATGGCAAC 7560

ACTTGTAACT TTTTTAGCTA CTGTCAAGAC AATTTGAGTA GGTTTACTCA CATCATAAGT 7620

CGTTCCGGCA CCTGGACTTT GTTTCATAAT CGTTCCTGGT TCGCTTTCGC TGGACTCTTC 7680

TTCCTCTATC TTAATCAAAT TCTCAGGAAC CTTCTTCTGC TTGAGTTCTG AGATTACTTC 7740

TGTAGAGTTC CGTCCAATAT AGTTCCCTAA TTGAATCGTC GTAGCTTTTT TAGCTACTGT 7800

CAAAACAATT TGAGTTGCCT TGCTCAAGTC ATAGGTCGTA CCTTCTGGTA GACTTTGCTT 7860

CAGGACCGTT CCAGCCTCAC TCTCATTCGA CTCTTCTTCC TCAATTTTAA TCAAATTATC 7920

TGGAACTTTT TTCTCTTTTA ATTCCGCAAT GACATCAGAG GATTTCCGAC CGACATAATT 7980

ACTAATTTGG AAAGATTGCT TGCCTGATGA GACAACCAAA TTGATTTTCG TTCCTTCTTT 8040

TCGACCAGTT CCAGCGCCAG GATCTGTACG GATAATCCGC CCTTCTTCCA CCTTTTCACT 8100

AGCCTCTGTC TTCTCCTCAC CAATCTCAAA ATTGGCTTTT TTGAGCGTTG CCTTGGCCTC 8160

TGCAACTGTC TGACCTGCCA CATCTGGAAT GGCAATGGTT GCAGGAGTTC TGGATAGTAT 8220

CCAAATAAGA GAAGCTGCCA CCAATACAAG GCTGGCCAAC AAAATCAGGT AACGCATCTT 8280

AAATCTATGT TTTTTCGGTG CTTGTGGTTG GTAAGTTTCC TCTGTCACAG CCTGGCTTGG 8340

GTTTTTGATT GATTTGTGTT CTGTTTGCGC TTGAACCTTA GGAATAGATG TCAAGGTACT 8400

CTGAGAAACC TTCGGCAAGG TCTTGGTATC TGCCTTGCTC GTTTCATCAA AGATTAACTT 8460

ACTTTCATTT CTACGATTGT AGGACAAGCT ACTAGACAAG TCCACATACA TCTCTGAAAC 8520

CGAGCGGTAG CGATTGGTCA ACTTTTTAGC AGTTGCCTTG ATAATAACAT TTTCTAAAGC 8580

CTGAGGTACA GATGGATTTT CTGCAATAAC GGACGGCAGG GGTTTCTGGA AATGCTGGAG 8640

GGCAATGGTC ACCGCGCTAT CCCCGTCATA AGGGATATGG CCTGTCAGCA TCTCATAGAA 8700

AATAATCCCC ATGGCATAGA TATCACTCTG CACAGTCGCC TTCGAACCAC GCGCCTGCTC 8760

TGGTGACAAG TAATGAACTG AGCCCAACAT CGAGTTAGTC TGGGTCAGAC TTGTCTCTGC 8820

AAAGGCTACA GCAATCCCAA AGTCTGTGAC CTTGGCAGTC CCATCTGGTG TCAAGAGGAT 8880

ATTTTGAGGT TTCAAGTCCC TGTGAACAAT TCCTCGAGTA TGGGCCAAGC GCATAGCCAA 8940

GAGAATTTGT CCCATGATAC GGACTGCTTC TTCATTAGAA AGAGGATAAT GTTCCTTGAT 9000

ATAGCGTTTG AGGTCCAGTC CAGCCACATA CTCCATAGCT AGGTACTGTT GACCGTCTTC 9060

CTCGCCAATA TCTGTTATCC GAACGATATG AGGATGGTCT AGATCTGCCA TAGCTCTCGC 9120

TTCACGCTGA AAACGAGCTA CAGCTATCGG GTCCGTCTGG TAGTTGGTCC TCAGAACCTT 9180

CACTGCCACT TCTTCCCCAT CTAAGATTAA GTCTTTGGCT AGGTAGACAT CCGCCATACC 9240

TCCTCGACCA ATCTGTTTGA CAATCCGATA GCGTCCGGCA AAAATCTTGC CGATTTGGAT 9300

CATTCTGCAT CCTCCTCGTT CATAGAAACA AGGGCAACCG TAATGTTGTC TAAACCTCCT 9360

GCATTGTTAG CAAAACGAAC AAGTGTCTCC GTTTTATCTG CTAAAGGAAT ATCACTGGTT 9420

ACAATATCAC GAATCTCACT GCCTGAAATC ATGTTGGTCA AGCCGTCACT ATTGAGCAAG 9480

AGATAGTCAC CTGACTCAAG GATAACTGTC CCAAAATCAG GCTGAATTTC ATCTTTTTGC 9540

CCAATAGACT GGGTGATAAT ATTTTTTTGC GGATGAGCTT CTGCCTCTTC TGGTGTCAAT 9600

TGACCAGCCT TGAGCAATTC ATTAACCAAG GAATGATCGC TCGTCAACTG ATGGTATTCT 9660

TCTCCACGAA TCAAGCCGAT ACGCGAATCA CCAATATGAG CATAGATAGC CTGATTATCA 9720

ATAATAGCAA GGACTTCCAA AGTAGTTCCC ATGCCTCTGT AAGCTTCATC CTGACCAAGC 9780

TGGTGAATCT TTTGATTTTC AATTTCTAGG TAATGGGCGA ACCATTCACG CACTTCATTG 9840

ACTGTATCGA TCTGGGTATC AACCCAAGCT ACACCCAGGT CTGTGACCGC CATTTCACTA 9900

GCGATATTCC CTGCGCGATG ACCTCCCATC CCATCAGCTA AAATAATCAT GGTACGTCCA 9960

GCTCTATTGA CATAGTGGTT GACATAGTCT TGGTTATTTG TTCGTTTCTG ACCAACATCT 10020

GTTAATAATG AAATTTCCAT GTGTCAGTTC CTTCCTAATC CGATATCTTG CGAAATTGAC 10080

TGATGAAGAA TCCATCACTT CCATACAATT CAGGTGTAAT GAGGATACAG CCGTCTTTCA 10140

TGATATCCTT ACATTCATGT TCTAGTTTTA CCTGCTCGAA CTCGGGATGA CTCTCTAAAA 10200

AGGCCTTAAC GACTTGAAAA TTCTCCTCTG AGACGATAGT GCAGGTGCTA TAAGTTATTA 10260

TACCACCTTT GCCTAGTATT TGACAAACAC TACCTAATAT TTCTAACTGA ATTTCCTGCA 10320

AGGACGCGAA ATCTGCCGTT TCTTTATTGT ATTTGATATC TGGTTTTCGG CGCAAAAGAC 10380

CGATTCCTGA ACAAGGAGCA TCCACCAAAA TCTTATCAAA GGAATCCTGG TCAAAAAACT 10440

CATGCACCTT TCTGGCATCC AATTTTTGAG TTTGAACCCG ATCTTCAACT CCCAGACGTT 10500

GGGCATTTTC TTGAATTAAA TCCAACTTGT GGTCGTACAA GTCCAGAGCA GTAACCTGAC 10560

CTGTCGTAAG ATAAGAGGCT ATATGGGCTG TTTTCCCACC TGGAGCCGCA CAGGCATCAA 10620

GCACTCGCTC ATCACCTTGT AAATCAAGCG TCGGAGCAAC CAGCTGACTG GACTCATCTT 10680

GGATGGTAAT GGCTCCATCC GCAAACAAAT TATGCCCTGC AAAATGCCCC TGCTCCTTAA 10740

CCAGACCAGT GGTTGCTAAA AGGGAATTAT TCGCCTCCAA CAAGGCTTGG ATTTCCTCTT 10800

TTCGACTTAG GTCTGTTACA CGAATACTGG CTTTGTTTCG CACTAACAGG CTTTCAAAGA 10860

TGGCTTTTGC TCTCTCCTCT CCGTATTCTT CCTTGAGTTT GGCAACTAGC CAAACTGGGA 10920

GAGAATAGGC AATGGAGTCA CGCTTGTTTT TTCGCTTGAT GCTAGCAATA TCTGGCCAGC 10980

CTTCACGCAA GATACGGCGA AGGACAGCGT TGACCAATTT TTCACTGCCT TTTTTACGGA 11040

700

GTTTGGCCAA TTCCACTGCT TCATTAACCA CAGCATGATC TGGAATCTTG TCCAAATAGC 11100

GGAGTTGGTA GGCACTCATG AGAAGAAGGA CATAGAGCCA GCTGTCTAAC TGGTCTCTGT 11160

CTTCGATAAA GTGGGATAGG TACCATTCCA GAGTCAGTTT ACGGGCTACC GTTCCATAGA 11220

CCAGCTCGGT CACTAAGCCC TTGTCTGCTG CCAAAAGTTG ACTTCCCTTT AGATGCTTAT 11280

TTAAGGCGAT ATTTGAATAT GCTTGGTTCA CAAAAACATC CTCTAGCACT GCTAGAGCTA 11340

AACTTCTAGC CGTTTCTACT TTAGTCACCA AATCGTTCTC CTACAGTCAA TGTACGTCCA 11400

ACTCCGTTGA GGAAGGAAGC AATGTCCATC TTAGGCTTAC CAGCTGGCTG CACTTGTTTG 11460

AGGGATAGAG CCCCTTCAGC CGTTGCGACA ATCAATTCTT TCTTGCCGAT AGAGAGAATC 11520

TCACCTGGAT TTCCCTGACC TTCTACTGGT AGGGCTTCAT AAATCTTAAA GCGGTCGCCC 11580

TTAAGGAAAG TATGGGCAAC AGGCCAGGGG TTCATTCCAC GAATTTGGTT AAAGAGTTGA 11640

CGATTGGTTT TGTTCCAGTC CAGTTTTTCT TCCTCTGGCT TTATATTTGG AGAGAAGGTA 11700

ACCTGACTCG TATCCTGCGG TTCAGGTTTG ATATCACCAG CAATATAGGC AGGCAGAGTG 11760

TCCAAAAGCA AATCACGACC AACTAGCGCC AATTTTTCAA ACAAGGTGCC AACATTGTCC 11820

TCATCTGTGA TCGGAATGCT GCGACGAGAA ATCATATCTC CTGCATCCAT TTCCTTAACC 11880

ATTTCCATGA TGGTCACACC AGCTTCCTCA TCCCCTTGAA TCAAGGCATA ATGGATAGGC 11940

GCACCACCAC GGTGTCTAGG AAGGAGGGAG GCATGAACGT TGACAGCAAA GTCCATGCTA 12000

TCAAGGAGTT TGCTTGGGAG AAACTGCCCA AAAGCAGCAG TCACAATTCC ATCTGCTCCT 12060

AGCTTCATAA GATCTTCCAT CTCTGGACTT CCAGATAATT TTTCAGGTTG GTAGATAGAT 12120

AGTCCTGCTT CCTTGGCAGC CTGCTTGACT GGGGTTTCTT GGATAACTTT TTTACGACCA 12180

ACAGCACGGT CTGGCTGGGT CACAACGGCT AGAATTTCGT AACGGTCATC TGTCAAAAGT 12240

CCTTTTAAGA CTGTTGCTGA AAAGTCGGGG GTCCCCATAA AGATTAGTTT TGTCATATCT 12300

TCTCCTTCTT ATAAAAATTG CTGCGGCTCA TGGTCAATGC TGAGACGGAG CTCACTATTT 12360

TCCCGTTCTT GAGTCAAGGC TAAAACCTGG TTGAGGGTCG ACCCCAGCTC ATCTTCTAAA 12420

CGGTATTTAA TTAAAATCTG GTAATGATAG AGGTTGTGGG TACGGGCAAT CGGTTTTGGC 12480

GTTGGCCCCA GAATGGGACT GGTCTCTGAC AAGCCTGACC GCAAAATGTT CATGACTTCA 12540

TAGGCACGTT TGAAAACCTC TTCTTCTTTC TTGTGAGAAA GGGTAATACC AATCGTGAAA 12600

TAGTAAGGTG GATAGCCGAG TTGTCGTCTG ATTCCCATTT CATAGGCATA AAAGCCTTCG 12660

TAATCTTGAT CCTTGGCAAA TCGAATAGCA TAGTGCTGCG GATTGTAGGA CTGTATCAAG 12720

ACTTGACCTG CCTTTTCAGC CCGACCTGAT CGACCTGCCA CCTGAGTCAA GAGCTGGAAG 12780

GTTCTCTCAG AAGAACGGAA ATCAGGCAGA TTCAAGGCCG TATCCGCATT TAGAACTCCG 12840

ACTAGGGTAA CATTGGGAAA ATCCAAACCC TTTGCAATCA TCTGAGTACC AAGTAAAATA 12900

TCCGCTTCCC CTCGCCCAAA CTGGTCAAGC AAGGCTTGGT GACTGCCTTT CTTTCGAGTC 12960

GTATCCACAT CCATCCTCAG AATGCGAGCT TGGGGAAAGA GTTCTGCTAG CTCATCATAA 13020

GCCTTCTGAG TTCCCGTCCC ATAGTAACGA ATACTGCGGC TCTTACAGTT AGGACAGACC 13080

TGAGGAATAT CCTTCGAGAA ACCACAATAA TGGCAGTTCA TAGTCTTGGT ATCCATATGC 13140

AAGGTCAGAG AAATATCGCA GTTGGGACAA GTATCCACCG TCCCACACTC CCGACACATG 13200

ACAAAGCTAG AATAACCACG GCCATTGAGC ATGAGAACCA CCTGCTCTTT TTTAACCAGA 13260

CGGTCTTGGA TAGCCTCTAG CAAAGGAGGC GTAAAGTTTG ACGTCTCATT TTGTCCGATA 13320

TAGTCTCGAA AGTCAATCAC TTGAACCTCA GGGATTGTAG CCAAAGGATT GGCACGTTGG 13380

GTTAGACGTA AGTGTTGATA GACGCCTTTG CCAGCACGTG CCCGGCTCTC TAAGCTCGGC 13440

GTTGCAGATC CAAGTACCAG AGTTGCTTGA TTATACTGAG CCCGTAAAAT AGCTACCTCT 13500

CTGGCATGGT AACGGGGATT GCTGTCCTGC TTATAAGCCG CTTCATGCTC TTCATCAATA 13560

ATCATGACAC CCAGATTTTT CAGAGGAGCA AAGATAGCAG ATCTGGCACC AACAACAACT 13620

TGGGCATCGC CACGCTCCAC CTTGCGCCAT TCATCATACT TTTCACCATT GGATAATCCT 13680

GAGTGAAGAA TGGCTACCTT GTCCCCAAAA CGTGCTATAA AACGCTCGGT CATCTGAGGA 13740

GTCAAGGAAA TCTCAGGTAC CAGCAAAATA GCTGTCTTGC CCTTATCCAG GGCACCTTGG 13800

ATAATCTGCA AGTAAACCTC GGTCTTCCCA CTTCCTGTAA TCCCTTGAAG TAGAAAGGGA 13860

GGTTGAGAAC TGCCAATAGA ACTCACAACC GCATCACGCG CCTGTCTTTG TTCTGGATTT 13920

AACTCCAAAG GTCTACTTGC TTCAATTCCT TCAAAATAAG CAGCCGAGCG TTGAACTTCC 13980

TTTTGGACTA TGGTAACAGC ACCTTGATCC ACAAAGAAGT TGACTTGCTC TCGCGAGTAG 14040

GACTCTAACA AGCTAGCCAA GGAAGCGCTC TCTGGATGAG ACAGCAGATA ATCTCTCAGT 14100

TCCAACTTTT TCTTGGCACG TGTAGAAATC TCAACACCTT CTAATTGAGC ATGGTCAACC 14160

TCATACCAAG ACTGGGTCTT GACCTTCTTT TGATCGACTC CCTGATATTC CAGACCAAGC 14220

AGGCCTTTTC TAGTCAAACG CATCATTTCA GCTTGCTTGG CAAGGTCTAG TGAAGAAAAG 14280

GCTAGCGAAT CTTCTGAACC AAACAGGCGC ACTCGTTCTT CCTGACTCAA GCCTTCCAGA 14340

GGATAGAGAA TCTTGTCATA GCTAGAATTC AGAAACCCTG GAAGCATGGC CTTGAGGATA 14400

GAGATTTTGT AGGAGAAGAC AGATTTGCGT AACTCCTCAG CCAGCCAGAG TTGTTCTGGC 14460

GTGAGAACAG GAGAAAAATC CAGCACCTCT GCAATATCTT TTAAATCTTG CTCCATCTCT 14520

TCTCCATCTG ATTGGGACTT CAAACCAAGA ACAATCCCTT GAATCAGGCG ATTACCCTTA 14580

CCAAAAGGCA CATGAACCCG CATCCCAACT TCCAGCATTC CCTCAAATTC CTCCGGAATC 14640

CTGTAACTAT AGGGCTGGTC CGTCTGCATC AAGGGCACAT CTACGATAAT CTTAGCTAGG 14700

GCCATCTTCT CACCTCCTCC TTGTCAGTAC ATTCTTGCAA TAGAAAAAAT AAGATTGAGT 14760

CCCCCCAACC TTAAATTTTT TCACCATCTT CTTTTTCTTT AGCAATTTGC TCTTTGATTT 14820

TCTTTTCTTC TTCTTCTTTG CGGCGTTTTT CTTCTTCGAT ACGGCGACGC ACTGCTTCAC 14880

GTTTTCCTTC TGGATCTGGG TGAATTGTAA CGTTTCCTGA TTCGATTTCT TCTAAAGCGC 14940

GAAGAGTTGA TTTTTCAGAC TTGAAACCTT GAGTTGCTGG GGCACCTGCT TCCAATTCGT 15000

GGGCACGTTT TGCTTCCAAG ATTACGAGTG AATATTTTGA AGGAACCTTG TCGAGCAAGG 15060

TATCAATAGA GGGTTTTAAC ATCATTTGCT TGTACCTATT TTCTAAATTT TATCGGGTAG 15120

TTGGAGATTT TGGTAACATC TCCTGATAGT GACCAATGAC ACGATCCACA CAGAAGTGTT 15180

CTGCTTCAAT CACACATTTG ACACGTTCAG CAGCTAGGGG TACCTGATCG TTGACAATCG 15240

CATAATCATA CTCACGCATG AGGGCAATTT CTTCCTTGGC CTTTTCGATT CGTTGGGCAA 15300

TCACTTCTGC ACTATCTGTT CCACGACCTA CCAAGCGATC TTGCAATTCA TCCAAATCTG 15360

GTGGTGTCAG GAAGATAAAG ACAGCATCTG GAACCTTTTT CTTGACCTGA AGAGCACCCT 15420

GAACTTCAAT TTCAAGGAAA ACATCGATTC CCTTGTCCAA GGTTTCATTG ACATAGGTCA 15480

GAGGAGTTCC ATAGTAGTTA CCGACATATT CTGCGTATTC CAACATCTGT CCTTGACGAA 15540

TCAGCTCTTC AAATTCTTCA CGAGTACGGA AGAAATAGTC AACACCGTCC ACTTCTCCAG 15600

GACGTTGTGC GCGTGTCGTC ATCGATACAG AATATTGAAA TTGGTTTTCA GAACTCTCAA 15660

AAATCTCTCT TCTAACCGTT CCTTTTCCAA CCCCTGAAGG ACCAGAAAAA ACGATTAGTA 15720

AGCCTCGGTC TGCCATTGTG TCTCCTTTTA GTCAATCTGT GAAATAACAT TTCTCTAGAA 15780

TAATGGCAAA AAGCCAGATT ATCCTTTACA GTCTTTCTAT CTAGTGTAAC AAAAAAGCAG 15840

TAATTTTTCA ACTGCTCTTT CTTATTTATT TAGCATAATC TACTGCACGA AGCTCGCGAA 15900

TCACGGTTAC CTTGATATTT CCTGGATAAT CGAGATTGTT TTCAATTTTC TTACGAACTT 15960

TGTGAGCCAA GATTGTGACT TTGTCGTCCT TGATTTTTCC TGGATTGACC ATGATACGAA 16020

TTTCACGTCC TGCTTGAAGG GCAAAGCTAG TTTGCACTCC TTCAAAGCCG TTAGCAATTT 16080

CTTCCAAATC ATGGAGACGC TTGATGTAGC TTTCAAGAGA CTCACTACGA GCACCTGGAC 16140

GGGCTGCGCT CAAGGCATCT GCTGCAGCGA CGATAACTGC TATCACGCTC TCAGCTTCAA 16200

CATCTCCGTG GTGACTAGCA ATCGTATTCA CCACAACTGG GGGTTCCTTG TACTTACGGG 16260

CCAATTCCAT ACCGATTTCA ACGTGGCTAC CTTCAACCTC ATGGTCAATG GCTTTCCCGA 16320

TATCGTGAAG GAATCCAGCA CGACGGGCAA GAGCCGCATT TTCACCAAGT TCGCTCGCCA 16380

TGATACCAGC CAACTTAGCA ACCTCAATCG AATGGCGCAA AACATTTTGT CCATATGAAG 16440

TACGGAACTG CAAACGTCCC ATAATCTTCA TCAAGTCTGG ATGAAGGTTT GGCGCACCAA 16500

TTTCATAGGC AGCAGCCTCA CCGTATTCAC GAATCTTATT GTCAATCTCT TGACGGTTTT 16560

TCTCAACCAA CTCTTCGATA CGAGCTGGAT GTATACGACC ATCTTTGAGC AACATTTCCA 16620

TAGTCATACG GGCAATCTCA CGACGAATCG GATCAAATCC TGACAAGGTC ACCACTTCTG 16680

GTGTATCGTC GATAATCACA TCGACCCCTG TCAAACTTTC AAAGGTACGA ATGTTACGAC 16740

CTTCACGACC AATAATGCGT CCCTTCATAG TATCGTCTGG CAGATGAACT GTTGAGTTTG 16800

TTGACTCCGC TACATATTCA CCAGCGATAC GTTGCATAGC TTGAACCAAG ATGTCCTTGG 16860

CCATTTTGTC AGAACGTTCC TTGACCTCTT GCTCAGCTTC GCGAATGCGA CTGGCAATCT 16920

CCCTGGTCAA GTTTTCCTCT GTCTGAGCCA AGATAATATC TCGTGCTTCT GCCTGAGACA 16980

GCGCACCAAT ACGCTCTAGT TCTGCTTCTT TTTGTCTTTC GACTTCCTCT AATTGCTCTT 17040

CACGCGCATC AAGGTTTTTC GCTCTATCAG AAATACTTTG TTCTTTTTGT TCAAGTGTTT 17100

GTTCTTTACT CGTCAAATTG TCGTCCTTAC GGTCAAGGCT AGTAGCTCTC TCTGTCAAAC 17160

GACTTTCGAT TTGTTTGAGT TCTTGACGTT CTGATTTGAA TTCAGCGTCC ACTTCTTCAC 17220

GGTATTTTCT GGCTTCTTCT TTGGCCTCCA ATAGTGCTTC TTTTTTAAGA GACTTGCTTT 17280

CACGTTTGGC TTCATTAACA AGTAAATCCG CTTCACGCTC AGCTTGTCCA CGTAAATTAG 17340

TTGCTTCTTG TTCAGCATTT AAAAGCATCA ACTCTGCAGC TTCCTGAGAT GATTTCATCT 17400

TAGCTGAGAT GCTGACATAT CCAATGACTA AACCAATGAT GACGGCAAAA ACAGCAATCG 17460

CAAGCGACAT GATTTCCATG TTTTTACCTC ATTTTATTGT TATTCCGAAT GACATACATT 17520

CTTTTACATT CTACCATAAA AAAGTGATTT TCACAAACCT AAAATAGAAT ATGTTTTGAG 17580

GAATTTGGAA CACATTTACC AAAATAAACT TGTTGTTTAG AAATAGTAGT TTAGTAGAGA 17640

CTTGAGAAAA AGCCTACCTT TCAATAGACT TAGTAATGAT CTTTAAAGGA CAAGAAAGCC 17700

ACGCTATCTC CATCCATCAT ATAAATCAAG CGATTTTCTG CATCAATACG CCGTGACCAG 17760

GCTCCTTGGT AATCATATTT GAGTGGTTCT GGTTTACCTA TTCCTGTAAA GGGATCACGT 17820

TGAATATCCT TGATTAGTTT ATTGATTCTT TTTAACGTTT TCTTATCCTG ATTTTGCCAG 17880

TAGCAATAAT CTGCCCAGGC ATCTTCTGTA AACTTGAGCA GCATTTCTTA CTCCTCAATA 17940

ACATGGACCT GAGTACTTCC AGCACGAACT TGAGCCATTC CTCGCAAAAC CTTATCAGAA 18000

AGTTCCTTAT TTTGAGCAAT TCTCAGGGTT TCTTGGATAC TATCCCACTC ACTCTTTGAA 18060

AGGACTACAA TGTCCTCATC TGGATTTTTA TTGACCACCG TCAAAGGCTC AAATTCATCA 18120

TTTACCTTCT TCATGTAGTC CTTTAAATGA TTTCGGAATG TTGAGTAAAG GACTGCTTCC 18180

ATAACCATAC CTCGTTTTAG CTCTTTTCCA CTATTATACA CGAAAAGAAA GAAATTGTCA 18240

GGAACTTGTA CAAGATTTTC TTTTCTATCT ATTTATACTC AATGAAAATC AAAGAGCAAA 18300

CTAGGAAACT AGCCGCAGGC TGTACTTGAG TACGGCAAGG CGACGTTGAC GCGATTTGAA 18360

TTTGATTTTC GAAGAGTATT ATTCGTAAAA AATCTCAAAA AGCCTACCTT TCGGTAGACT 18420

TAGTTTGTTT CTATTC 18436

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7001 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

ACGTAGAAAA ACTATTTCTA TCACAGATAA TATTCCGTAT GTTGTTGGAG GTATTGAAAT 60

AAACGTCCTA GGTATCTTTC TCAGTCTATG TGACTTACAA GGGAAAACTC TTTTCGAGAC 120

AGAAATTTTG AATGAAGATT ATCCTATTTC AGAAATCAAT TCCACCATTA CCAATATGAT 180

AAAAACAGCT ATAGAGTACG TCCCTTTGGA AACAAAATTA CTTGGATTTG GCTTATCAAT 240

ACCTGGACAT TATAACAAAG ACTCCGGAAG TATCATTACA AACAACCCCA TATGGGAATC 300

TTTTAATTTA TTAAATGTAA TTAAAAGATT CAATTTTCCT TTTATTGTAA AAAATAATAT 360

CGATTGTATG GCTATAGGAC AATACCTTTT TAATCCACAC AATACCCCCG ATAACTTTAT 420

TTTCCTACAC GCTGGATTAG GTATTTACAC TTCCTTTTTC ACAAAAGAAA AAATAGGAGC 480

CTCTAAAAAT CCTTATATCG GAGAAATTGG ACACACCATT GTCGAATTGA ATGGGCAATA 540

TTGTGAATGC GGAAAAAAAG GTTGTTTACA AACATATATT TCGGATGCTT GGTTAATCAA 600

ACACGCCCAA TTATTATTTA AAAATTCCCA ACTAACTGTA CTAAAAAGCC TTGTAAAGAC 660

TGAAAAAGAC ATTCATTTAG ACACCCTTTT AACGGCTTAT AATTTAGGCG ACTCCGCTTT 720

ACGTCAACAA ATTGATAAAG GAGTCAATTT ATTAGCCACT TCTATTGCAA ATCTCCTCCT 780

CATCAATCCT GCTGATAAAA TCTATATCAA CAGTCAATTG CTTAATTATC AACCTTTCAC 840

TCATGAAGTC AGGGATAAAA TCCAAGACCA GCTCCACTTC GTTCCCTTTA CTCGTAATAT 900

AGAAATTGAA ATTTTACCTT ACAACAAACA TCGTGGAAGT ATAGGAGCTT GTGCATTAGC 960

TATCGTCGCT TTTTTCATAG AACATAGCAA TGTATTACAA GATATTATTT CACCTTAATA 1020

TATTAGAAAT CTATAGACCT GTTTAAATCA ACTATAACCT GTAGTAGATA TCTCGTATTT 1080

AGACAATATG AAAACAAGAC GACTTCCATA TAGGAAACCG CCTTCTCGCT ATGTTGAGTG 1140

ATTTATATTA AAATAACTTT TCTTCTAGCT GCATTTTATT ATTATAAAAA CATTCATCAT 1200

AACCCCCAGA ACTTAAATAA CAATTTTTAT TCAAGATACA TACTCCTAGA ATAAACTTTA 1260

TATGAAATTC TCATTTTTGT TTTTACAATT CTCCTTAGTT AAATCTTGTT TAATATATGT 1320

TTTACATATA GTATTTAGCG CCACATAGTA CTGAACTCTC TCCAAAAACG GTTATTCCTC 1380

TTTGAATAGG GCGTTATCAC AAGAAAAGCA TCTCCACGTT TCAACTTCAT ATGGCTCAAA 1440

AACAATCAAT TGATGCTAAA ACCTGTACCT AGATGTTTCG GTTCATAAAA CCATGAAACT 1500

GTAAAAGTGG ATGAAATTGA TAGCGATAGT CAAATCAAGA GGCATCATAA CTCTAAAAAG 1560

TCACAATATA TAAGTTCATC CTCGGAAAAA TATCATTCTA ATTGTTGAAA TGCCTACATG 1620

AAAAGAAACG TCAAATGCTC ATGAAACAAC GAATACAGGT ATCAAAACTA TGACAAAACA 1680

AATCCCTAAA TTTACTAAAG ACACTGCTCA ACTTTACACC TGTAAATGGT TGTTGTATAA 1740

TAAAGTTACA AAGATGTACG ACCACACTGT TGTAAATCAT AGTGTTCGCG AATATATTAC 1800

TGATAGCATT TCTACAAATA CAAGTAAAGA GAGCGGATGA GATTCAAACG AAATATGTCA 1860

GTGCTTTGGC ATTCCTAGCC TTCATATCAT TTAAAGAATT GTATAGACAA AATTTTTTCC 1920

AATACAGACA CTCGTAACAA CTGCTTCATT TTTCTACCAA CATATTTAGG AACAGGATAA 1980

GATACAAGAG TATTAATCCA TAGCTCAGTT CTATACCAAT CTAAGACAAA TAAGCTAAAA 2040

AAACGATTGA TAATAAGCAA ATAGATTCCA AATTTTCTCT ATCTGCTCAT TTTAATAAAC 2100

AATACTAGTG TAACTATCCT TCCAGTCAGA AGCTTGTCAA ATCACACCGA AAATTCTTCT 2160

AAAATTTATC TCGTTAGGCA ATCAAGCAAA AACTCGACGA TAGTACAAAC ATTATCATAC 2220

AGGATTGACT TCCTAAATTA TATACTTTAG TAAGGTTTTC GGATAAGAAA AAAGGTTCAT 2280

TTTACATTTC TAAACATTCT TTTCTAAGAT GAAAAACAGA ATTTTTCGAT TGTGATTTAA 2340

AGCAACAAGA AGATTTTCAG TATCATCCTA TAGATACGAG CTAATTAAGA AAAACTACAT 2400

TTTTGAATAT AAACTACAAT AATATAAACT AAATTTTATA GGAGGAAGAC AATGGATTGG 2460

TACGATTATA TGATACAGGC ATCCAAACAA TCACAATTCA ACGCAAGCCA TTGGTTTCGC 2520

TATTTGCGAA AAGTTATTTT TGAAGACTAT TCTTATTTAA CAAACCAAGA TGTAGAAAAG 2580

TTGCTAGACT CCAAAGAACT AACCCGTTTT CAAAAAATTA GCTTGAAGTA TGCCTTTCAA 2640

GAGCATACTC CAACTCATAA ATATGTGATT TCATTAAATA AACCTGCTAA GTTAACCAAT 2700

GTTCAAAAAT TGATGGAGAA ATACAAACAT GGATAAAATG AAACCGGTCT TCCAAGCCCT 2760

AAATAAGGAA TTAATTCAGG AAAATCTGAC TTTAACAATT ATCTGTGTCG GTGGTTATGT 2820

CTTAGAATAT CATGGTTTAC GTGCCACACA AGATGTTGAT GCTTTTATGG CTCTATAATA 2880

TTTGTAGTGG GTAAATCCCC TATGGATATT ATGGAGCCTA TTTTTGTGTA GAAAAAAAGT 2940

CCCATATGAC CTATAATGAA AAGCGACAAA ACAACTCATT AGAAAGAATC ATATGGAACA 3000

ATTACATTTT ATCACAAAAT TACTAGACAT TAAAGACCCT AATATCCAGA TTTTAGACAT 3060

CATCAATAAG GATACACACA AGGAAATCAT CGCCAAACTG GACTACGACG CCCCATCTTG 3120

CCCTGAGTGC GGAAACCAAT TGAAGAAATA TGACTTTCAA AAACCGTCTA AGATCCCTTA 3180

CCTCGAAACA ACTGGTATGC CTTCTAGAAT TCTCCTTAGA AAACGCCGTT TCAAGTGCTA 3240

TCACTGTTCA AAAATGATGG TCGCTGAAAC TTCTATCGTC AAGAAGAATC ATCAAATTCC 3300

TCGTATTATC AACCAAAAAA TTGCGCAAAA GTTGATTGAG AAGATTTCTA TGACCGATAT 3360

TGCTCATCAG CTGGCCATTT CAACTTCAAC TGTCATTCGC AAGCTCAATG ATTCTCACTT 3420

TGAGCATGAT TTTTCGCGTC TTCCTGAGAT TATGTCCTGG GACGTTGAAA CAGTCCGGGG 3480

AGTGACTGTT TCAATCGGGA GATGGAGATG AGCTTTATTG CGCAAGATTT TGAAAAGCTC 3540

GATATCATCA CTGTTCTTGA AGGTAGAACA CAAGCTGTCA TCCGAGATCA CTTTCTTAAA 3600

TATGATAGAG CCGTCCGATG TCGCGTCAAA ATTATTACTA TGGATATGTT TAGTCCTTAC 3660

TATGACTTAG CTAGACAACT TTTCCCGTGT GCTAAAATCG TTCTTGATCG CTTTCACATT 3720

GTACAACATC TTAGCCGTGC TATGAGTCGT GTGCGTGTCC AAATCATGAA TCAGTTTCAT 3780

CGAAAATCCC ATGAATACAA GGCTATCAAG CGCTACTGGA AACTCATTCA ACAGGATAGC 3840

CGTAAACTCA GCGATAAACA TTTTTATCGC CCTACTTTTC GTATGCATTT AACCAATAAA 3900

GAGATTTTAG ACAAGCTTTT GAGCTATTCA CAAGACTTGA AACATCACTA TCAGCTCTAT 3960

CAACTCTTGC TGTTTCACTT TCAGAATAAG GAACCGGAGA AATTTTTCGA ACTTATCGAG 4020

GACAATCTTA AGCAGGTTCA TCCTATTTTT CAGACTGTCT TTAAAACCTT CCTCAAAGAT 4080

AAAGAAAAGG TTATCAACGC CCTTCAACTA CACTATTCTA ATGCCAAACT GGAAGCGACC 4140

AATAATCTCA TCAAACTTAT CAAGCGCAAT GCCTTTGGTT TTCGAAACTT TGAAAACTTC 4200

AAAAAACGGA TTTTTATCGC TCTGAATATC AAAAAAGAAA GGACAAAATT TGTCCTTTCT 4260

CGAGCTTAGC TTTTTTTCAA CCCACTACAG TTGACAAAGA GCCGGAAAAA GGAACAGCCT 4320

TAGCTTTCCT TTCATTTCTT TTTATTTCCC TCGTAGTAAA CGTGCTAGCT TCCACAAAAC 4380

AAACAGGATT CCCAGAAATG CCAGTACCAC TAGCCCACGG TACAACCATT GAGAGGTTGC 4440

AACACGCGAT ACAGATTGTC CTTCTTTCGT AAAAGCAACC CTCGCAACTG CAGCTGTTTG 4500

TGGATCTGAT TTTTGATAAA CAGCGACTCG TTCAAAATTC ACTAATAAGC GTTTATTAAA 4560

GGTAGGAATC GGATCGCAGG TTATCAAGGT CATGATATTT TTAGAGCTAA CCGATTCTAA 4620

TTTTTCCCAT TCCGACGGTA AAATAATCTC TGTGTCCATC ATCTGATATT CTACAATTTC 4680

CTGGCCATTA TCATAATAAA GAGCATCTCC AACTTTTAGC TGATCCAAAT GGCGGAAAAA 4740

GACATGGCTT GGCTCTGCAC GGTGCCCAGC AATCACTGAG CGAATCCCTG TACCATCCAG 4800

AGGCAGCGGT GTACCATCCA CATGAGCCAA GCCCATCCCT AAATGATGAT AATCTGCTCC 4860

CAAATAAACC GGCTCCATGA TTTCCAAACT TGGAATAGAC AAGTAACCAT AGACTGCATC 4920

AGGGTCGTCA GACACTTGGT AATTGACCTC ATATCCCTCC GCCAAAAAAG GATCTACAAT 4980

GCGATTTTGC GAAGCCAAGC GTTGATTGTA GGCGAGAGAA TGGTTCTGTT GTTCTTGGTA 5040

CATTTCAGTT GTCATGGATT TCACAAATGT AGCATGACCT TTCACCTGTC CAAGAGACTG 5100

CAACACCATC TGTCCAAAAC AATAAATAGG AATCAAACAG GCTACCAACA TCAACAAGTA 5160

TCCCAATAAG GCTCGTAGTT TAGTCCTTGA CATGACGCCC CTCCAATTGC TTTTCTAGTC 5220

CTTTGACAAT CCGTCGATTA CGATACACGC GATACAGCAA GAGAAGGATG ACCGCCATCG 5280

CTCCTAGTAA TAACCACAAC CAGAATTGCC CACGCTCTCT CACCGCTCGA TTCCGCTCTG 5340

CAATTGGTGC CGTATACGGA ATCCGCTTCC CACGTACCAA CAGACGATGA CTGTTAATCA 5400

TATACGGTGT ACAAGTCAAC AAGGTCGCAT AATCTTCCCC ATGTTGAATC AAGACAGGCT 5460

CAAAGTCATT CGGCTCCACC GTCACTATCT GATCCACTTG GTAGGCCAAC ACCTGATCTA 5520

AAACGTGAAG ATAAAAGATA TCCCCTTTTT TCATCTTATC CAATTGACTG AACAATTCTG 5580

CCGTTGGCAA TCCTCTGTGA GCAGTGATCA CTGTATGGGT ATTTTCACCT CCAACAGGCA 5640

GCGAAGCCCC TTCTAACAGC CCTGCCCCTT TCTGAAGAAT GTCCTCACTC GTTCCGACAT 5700

ACATCGGAAT TTCCTGATCA ATCGCAGGAA TTTCCACATA GCCAATCCGC TCATGGACCT 5760

TTAGCATATT GGCATATTCT GAGACGCCTT TCTTTTTCTC TTGCTCTGTA AAAGGATCAA 5820

GAATTTCAGA TGGTTTCAAG GTCGCATTGA AGGCTTGAGC CAAGCGCCAA CGCTCCTCAA 5880

GTTCTGCCTT ATCCATCTGG GAAACCGTCT CATCAAACTC TTTAATAACC TCGTTTGACT 5940

CAATACGATA ATAATAACGA GACACCAATG GATATATCGC AACGGCGAAT CCTACTAAGA 6000

AAATCAGAAG AAGGATCAGC GGATGTTTCT TCTTTTTTGT GCCTTTTTTT CGTGAACGTC 6060

TACTGTTGTC CATCCTCCAC CTTCACTTCC TTCCTTGCTG CTTTCAGCGC CTTCAAAGCC 6120

TTTTCCGGTT GTTTTTTCTT CTTGCGCAAG CGTCGAATAA TCCATAAAAG AATCACAATC 6180

AAACCAACTG CCACATAAAA CAGGTAGCGA TAGAGATGAC TGAGTTTGTT TGCTGCAATA 6240

AATTCTTCCT CAACCTCTGC TACGTACGGT ATCCGATGCC CCCGAACCAA TAGACGATGG 6300

GTATTGATCA TGTATGGCGT ACAAGTCAGC AAGGTCACAT AATCATGACC TGGTACAATC 6360

AATAAATCAT CAAAGTTCGT CGGCTCAATC ACCTTTACTT GATCCACTTG ATAGGCCATC 6420

ACTTCCTTGA TATTGTGCAC ATAAAACTTA TCCCCAACTT TAAGTTTGGT CAAATCCGTA 6480

AACATCTTAG CTGTTGGCAA ACCTGTATGT GCCGTAATCA CCGCATGGGT CGAATTGCCT 6540

CCGATCGGCA GAGAAGTTCC CTCTAGATGC CCAGCCCCTT GCTGCAATAC CTCTTCAGCA 6600

GTACCAGCAT AAACCGGCAA ATCCACGTCA ATAACGGGGA TTTCCACATG CCCCATCCGC 6660

TCATGGATTT CTAACATACG TGCATACTCT GCTCGCCCTT TTTTCTTCAT TTCTTCCGAC 6720

CAAGGATCGC CACTCACTAC ATTATTCAAA GAGTCATTGA AGGCTTGTGC CAATTTCATT 6780

CGTTCATCAA TGTCAGCCTC ATCCAACGTT GCTTTTTCCT TATCAAAGTC AGCAATTTGT 6840

TGATTTGATT CCACTCGATA ATACAAGCGA GACACCAGCG GATACGCCAT TACCGCCATT 6900

CCAATGAAAA ATACCACTCC TAATAGGAGA TTATTTCGTT TTTGCTTTTT TGTTTTTACC 6960

ATTTTTATCA GCATCCCTTT ATCTTCAAAC TTCAGGGTAT C 7001

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10411 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

GAGGGAGCTT AAGAAGTTAC CACCGTCCTC TAGCGCCTTA TCCGCATCAA AGTTAAGGTT 60

GATATTTTTA AAACTGTCGC CAGCTTGTGA TACGATGCTT TGTTTAAGGT CATTTAGGGT 120

TTTAGTGAAA TCTGCATTGC TGAGGATATC ACTCTTTGAG AGATTCAAGG CAAAATTGAT 180

GATGATATTG ATCTGGTTTC CTGTTATGAC CTGATCAAGT TTGTAATTTT TTAAGGTATC 240

TTCAACAATC TTGCGGATAT CTTCTTCTGT CAGATTTCCC TTACTTTCTT TAGCTTTGGC 300

GAGTCCTGAC TTGATATCAG CTAGGGCAAC GTTTAATTTA TTAGCATCAT AGCCTGATCT 360

GTCCTTGTTT TCAGCATTGA TATCTGACAA AGCTTTTAGC TCTTCTTGAG CCAAATCTTT 420

ATTAGCTTGT GGCACCTTGG CTCCATTAGC CTCTAGCGAA TAGTAAATCC CTGCTAAAGC 480

ACTTTCTCCT GTAACTGGAA TAGGGGCTGC TACAGTGATT TTGGCATGTT CCATACCCAG 540

CGTTACTGCT GCGTTTCGGT ACATATCCTG AGTCACCTTA GTGATATTTT CTGGTGTTTC 600

AATCTTGACC TCAAGTGGCG ATTTGTCACC TAGCTTTTGA ATCTTGGCTG ATGAATACAA 660

CTGTAAGCTA GAGTCATTGG CCACATTCAT GATTTTAGAA TAAACATCAG GTGTCATGGT 720

CTTGAGTTCT TTGGTATCTG TTGAGGCATT GTAGCCCAGT TTTTTAAGAG TTTGATTTTT 780

TTGGTCTTCA GATAGGGAGG AACCTAGGAC ATATTCAGGT TGGACATAGG TTTCATCGAT 840

AACTTTTTGA ACATCTGTTG CTGCATGGAC GCTATTCATA GCTGTTACTG CCCACAAGAT 900

CGCAGCGCTA GTCAGAAAGA GTTTCTTTCT CATAGGGAAT TTCCTCCTTT ACTTCTTTAG 960

AGTAATATAT CTATCTTAAA GAAAACTTAT AACAAAAACA CCTGGTCTAG CCAGATGTTG 1020

AAAAGAGAGT GAAACATTTG ATGATGTAAA GGTTAAGTCG TACCTGTCTA GAATAATAAT 1080

AGTTTCCTCC ATTTACATAG AGTTCAGCAC CGTGAAAAAT GGAAATGGGG TGAATATAAC 1140

TATAAGTCTT TCCAGTCCTA TTACCAAGCA AGGGGGCAAC AGTCTCACGA GAGTACTGTT 1200

TGGCTAGAGC CAGGGTATTT TCCTTGCCAT TTTGGGCGAT AAAATCGATA TAGGCAGGTC 1260

CAAAATTATA GGCTTGAACA GCTGTCCAGA TATCTACCCC CTTCTTCTGC GCCAGATAGA 1320

GATTGCCTGT CAGAGTTTGA ATGCCTTGCC GAATGCTAGA GGCATTATCA TTGATGGTGT 1380

TGGTGGAACC ACTTGCAGAC TCACTAGACT GCATAACATC GCCTTCTTTT CCTTTTGTTT 1440

CAGTATAAAT CATAGCAAGC ACAAGCTCTT CGTTTGCTGG GGTGTCTTGT TCACTCAATA 1500

TTTCTCGCAC CATGGGTTGA TAGGTCATGA CTTGTTTGAC ATCTTGATGA ACGCGGTAAG 1560

CTTTATAGCC AGCAAAAAGG AAGACTGCTA GTACAAGCAC TCTTCGAATT CGTTTAAACA 1620

TTATTTACTT TGGATATCCT CGATATTTTT GATTAAGATA GAGTAGGTTC CATTTTCGTT 1680

TTGGATAAAC TCAACAGACT CGGCGTCTTG ATAGACGTTA TTGGGAACGA TGAGCTCAAT 1740

TCCATTTGAT AAGGAGAGTT TTTGGTTTTC AAATTTCTTT AATTGGCGAC TGGCATCAAT 1800

TTCATCAAAT TGAACAGGTT CTGGTACGGC TTCTTTGACT TGGTCAATAA AGCTCAAACG 1860

AGCCGTCAGA TTGTTGTCAA AAAGGTCATT AGCCAATTTC TCAGGTGACA ATTCATTGCT 1920

TTCTTCTAGG TTGTTGAAAA TAGCTGATTT GACCTTGGAT TGAAATTGAA AATCATCTGT 1980

GTTAAAAGAT TTAGCAATTC TCTGGGCTGT TTTTTCCAGT TCCTTGATAG ATTTTTTAGG 2040

AGAAATCTTA GGAGCGACAG CAAGAAGATT ATCTGAAAAA TAGTTCAAAA AAGTCCCGTT 2100

GTACTTGATT CGTTTTTCAA TCAGGTGATA CTTGCTACTC TGAAGATTGA CCACCAAGGC 2160

CTCATCAGCT CCTGTTCCAA ATCCAGGCAG GTTATTCTGA GTTAGCTTGA TTGGATTATC 2220

AACTTCTCCT CCGAGGTGGG TCAAGGTCTC CCGCAGGGCA ATTCGCAAGA AAGCGAAATG 2280

TTCTACACCT TCTTTAGAAA ATTGCACAAA AATCAAGTCA TTGGTCTTGA GATTTTCAGA 2340

AATGCTAAAC TCCTCTTTCC AGAGATTAGC CAGCGTTACT GATGTCTCCA ACAAATCGTC 2400

TGTAATATGA TTGAAGAAGG GATTTTCTTC TTCGAAAATC CCAGTCTTGG CTTCATCTGA 2460

ATACACATGT TCAATTTTTT TACGCAGGTA TTCTTCGATT TTTGGAGTAA TATTGAGAAA 2520

CTTATCTGCT AAGAACAGTT CGGTATCATC CGGACTGAAC TGGTGAATAA TGGCTTTCTT 2580

AATATAAATG TCCATAAAAG TTTTAGTCCT CGTATAATGG GAAGGCATCT GTCAATTCTT 2640

TGACTGCACT TCTCACTTCT TCTAATACAG CCTCATTTTC TGAATTCTTA AGGGTTTTAA 2700

TGATGAGTTC AGCCACTTTG CGACTTTCTT CTTCACCAAA TCCACGTGCA GTAATGGCTG 2760

CTGCTCCGAT ACGAATCCCA CTTGTCTTGA ATGGTGACAA GCTTTCGTAA GGGATTGAGT 2820

TTTTATTTAA GGTAATATTG ACTTCATCCA ACAAGTTTTG AGCAACTTTG CCGTTTTCTA 2880

CAACTTTAGT CACATCAACA AGGAAGAGAT GGTTTTCAGT TCCACCTGAA ATAATACGGA 2940

AATCAGGGTC TTGCAAGAAG ACATCTGCCA TAGCCTTGCT GTTCTTAATT ACATTGGCAG 3000

CATATTCCTT GAAGGCTGGA TCCAAAACTT CTTTGAAGGA AACTGCCTTA GCCGCCACAA 3060

CATGCTCTAA AGGACCGCCC TGAATACCTG GGAAAATAGC TGAATTGATT TTTTTAGCAA 3120

GTTCTTCGTC ATTGGTCAAA ATCAAACCAC CACGAGGTCC ACGAAGGGTT TTGTGGGTCG 3180

TTGTTGTTGT GATATGAGCG TATGGAACTG GGCTTGGATG AAGGCCAGCC GCAACCAAGC 3240

CAGCGATATG GGCCATGTCC ACCATGAGCT TCGCACCGAC AGCATCTGCG ATTTCACGGA 3300

ATTTTGAAAA ATCGATAATT TGAGAATAGG CTGAAGCACC AGCTACAATC AGTTTTGGTT 3360

TTACTTCTTG GGCTTGTTTC AAGATAGCAT CAAAGTCTAA GAGTTCCGTT TTAGGATCAA 3420

CACTATAAGA AACAAAGTTG TAGGTTTGAC CAGAGAAGCT AACAGGAGCC CCATGAGTCA 3480

AATGACCACC TGATGCCAAA TCCATTCCCA TAACCGTATC ACCTGGCTCA ATCAAGGACA 3540

TGTAAGCCGC ACAGTTAGCT TGGCTTCCTG AATGTGGTTG AACATTGGCA AATTTAGCAC 3600

CGAAAATTTC TTTTGCGCGT TCAATAGCAA GAGTCTCTAC AACGTCTACT ACATCAGTTC 3660

CACCATAATA ACGGCGTCCT GGGTAACCCT CGGCATATTT ATTTGTCAAG ATAGACCCTT 3720

GAGCTGCCAT AACAGCCTTG GAAACTACGT TTTCCGAAGC AATTAACTCG ATATTATTTT 3780

GTTGGCGTTC TTCTTCTTTG GCAATAGCAT TCCAGAGATC AGCATCATAT GCTTTAAAAT 3840

CATCTTTGTC AAAAATCATA GGTCTTCTCC TTTATTGTGT GACTAGTCCA TTAGTTTGAT 3900

TTTACAATAA GAAAATCAAA CTAACAGATG CGAATAAACC GTTTCTGCAT TTTATCACAA 3960

GTATAGCCAA CTTTTTCATA AAATGCATGA GCACCCAGAC GATGATTGGC AGAATTTAAG 4020

CGGATAAACC CATAACCACA TCTTTTTGCT TCTTCTTCCA ACCCTTGTAG TAAACTTTTA 4080

CCAATACCTT GACCTTGCGC TTGAGGTGAA ACTGCTAAAG CTAAGATATT AAATCCTGCT 4140

TTGGAATAGA GTGATTCGTA AACTTCAGCG TGGACATATC CAAGTAAGAC ATGATTAGCT 4200

GCATCCTCAT AGCCAAGTAG GAAATGATGG GAATCCTGAG ACAGTCTAGC TAGTTGGCTA 4260

GCCGTTTCCT CTGGACTAAA AGTATAACCC AAAGCCTCTT GGTTGATGTC ACATATAGCT 4320

TTCACATCAG TTTCTCTTAA ATCTCTTAGC ATCTCATTCC TCCTCAAAAG AAATCTTTGG 4380

CAACCGAGCA AGAATATCTT CTCGCTTAAT GGCCCCTTGA CGTAAGATTT TCACCTTGTC 4440

TCCCGACAAA TTCAAAATAG TTGAATCCTG TCCAGTTAGA AAAGCATCGT CTTCCAGACC 4500

CAGAACCTCT TGGTCAAAAT CCTCTAGAAT TTGATTAAAG GTCACTCCAC TCGCCTGACC 4560

TGAGATATTG GCAGACGGCC CAATCAAGGG ACCTGTCTCT CGAATCAAAT CAAGGGTAAT 4620

GGGATGACTA GGCATCCGAA ATCCAACAGT TGCAAGGCCA GAATTGACCC AATAGGGAAC 4680

TCGGTCATTA GCTTCGAGAA TAATGGTCAA GGGACCTGGT AAAAAGATCT CTACAAGTTT 4740

TTGAAGATAA GTTGGCTGAT TCTTTGAAAA GTACAAGATG TCCTCTAAAG AGGCAACATT 4800

GAGATTGAGC GCCTTGTCTC TACGTCGACG TTTAAGCTGG TAAACATGGT CAACTGCTTT 4860

TTCGTCTAGC GCCTTAGCAA AGAGACCGTA AACTGTCTCT GTAGGCAAAA CGACAGCTCC 4920

ACCATTTTCC AACTCTTGTC TAATCCTGTC CATCATCAAC GACAACCATC CTATCTTGAC 4980

CAAATTGGTC CTTGAGTGTT CGTACTCGCT TTTCAGGAAG ATGTTTCCTA AAAAGTTCAG 5040

GAACACTTTG ACCTTGCTTG TATCCAATTT CAAGGTAAAT CTTACCACCA TCTTTGAGAT 5100

AGTCTTTTGC ATCTTCCGCA ATTCTACGGT AAATAGCTAG GCCATCCTCA TCTGCAAAGA 5160

GAGCTAGATG AGGCTCCGAA TACAAGACAT TCAAGCCTAC CTCTGACTCA TCTTCACGAG 5220

AGATATAGGG TGGATTGGAA ACAATTATAT CATATTTTTC AGAAATTTCT GTAAAACAGT 5280

CAGATTTTTT TAAAAATATT TGAAGATTTT GATTTTTAGC ATTTTCGCTA GCTACATCTA 5340

AAGCATCTTG GGAAATATCT GCTGCCGTCA CTGACCAATC TGGTCTGTTT TTTGCTAGAG 5400

CGAGAGCAAT AGCTCCACTA CCTGTTCCGA TATCTAGGAC CATAAGATTT TTCACAGGAT 5460

TTTCAGCCAG GATAAGCTCC ACCAACTCCT CTGTTTCTGG ACGAGGAATC AAAACCCGTT 5520

CATCCACCTT TAAATGCATT CCATAAAAAT CTGCCTGTCC AATGATGTAC TGAGCTGGCT 5580

TGTGAGCTGC TAGTTGCTGG TAAATATCTT CTACAAATTG TTTTTCTTCC TCTGTTGTCA 5640

CCTCCTGCTG GAGGGCAAAA ATAAAGTCTG TAAAAGATAG ATTTTTCAGA CTACGATAGA 5700

CAAAAGAGAG GCTTTCCGCT TCCTCTCCTT GTCTTATCAA CTCTTCTTCA AAATTTGAAA 5760

ATAATTGAGC TAATTTCATT ATTTGTTTAA TTCTTCTAGT TTTTGTGTTT GGTCATAAAG 5820

CACCAAGGCA TCCACAACTT CGTCCAATTT ACCAGACAAA ATCGTATCTA GTTTTTGGAG 5880

GGTCAAGCCG ATACGGTGGT CTGTGACACG GTTTTGTGGG AAGTTATAAG TTCGGATCCG 5940

TTCTGAACGG TCACCAGTAC CGATTGTCGA CTTACGCTCA GCGTCCTGCT CATCTTGAGC 6000

AATCTGAGCA AAGTGGTCAG CAACACGGGC ACGGATGATT TTCATGGCCT TCTCACGGTT 6060

CTTCTGCTGG GTACGTTCTT CCTGCATCTC AACCTTGATA TTGGTTGGCA AGTGAACGAT 6120

ACGAACGGCA GTCGCAACCT TATTGACGTT CTGTCCACCA GCACCAGAGG CGTGATAGAT 6180

GTCGACACGA AGGTCTTTTG GATCAATGTC GTATTCAACC TCTTCAACTT CTGGCATAAC 6240

AAGAACTGTC GCTGTCGAAG TATGAACACG GCCTTGGCTT TCTGTCACAG GAACACGTTG 6300

CACACGGTGG GCACCTGATT CATACTTAAG CTTAGAGTAT ACAGACTGAC CTGAAACCAT 6360

AGCAACCACT TCTTTAAAAC CACCGACACC ATTCATAGAG GCTTCCATGA CTTCAAAGCG 6420

CCAACCTTGG GCTTCCGCAT ACTTTTGGTA CATAGTTAGC AAATCTCCAG CGAAAAGTGC 6480

CGCTTCGTCT CCACCAGCTG CTCCACGGAT TTCAAGGATG ATATTCTTGT CATCGTTTGG 6540

ATCCTTTGGA AGGAGCAAAA TTTTCAGTTT TTCTTCATAT TCTTCTTTTT CAGCCTTGGC 6600

ATCTTTGAGT TCTTGCTTGG CCAATTCTTC CAAGTCCGCA TCTCCGCCTG ATTCCTTAAT 6660

CATCTCTTCG GCATCGACGA TATTTTGAAG GACTTGTTTA TACTCACGGT AGGCTATTAC 6720

GGTGTCACGA TTGGAAGCTT CTTCTTTTGA AAGCTCCATA AAACGCTTGG TGTCTGAAAC 6780

GACATCAGGG TCACTCAGCA ATTCTCCTAA TTCTTCATAA CGGTCTTCTA CAACTTGTAG 6840

TTGATCATAG ATGTTCATTT TTTCTCCTTA TTTCTCAATT GTTAAATCAT AGATTGCTAC 6900

TACTTCATTC TCGGATATTT CCCCAGTTTC TTTAAATCCA TAACTGAGGT AACAAAATCT 6960

TGCCTGTTCA TTTTCTGGTT CATArGACAA CCAAAGTTTA TTGCTTAAAC CTGCTGGCGC 7020

TGTTCGAACA TAGTCTAGTA CTTTATCCAT AATTGGTTTA AAATATCCTT GATTTTGAAA 7080

ATTCTTATCA ATCATAAAAC GAAATAGTAA ATAATTTCCA CTACTAATTC CGATCTTTTT 7140

ATCATAAGCT ATCATCACAA AACCTATAAT TGCATCATTA TCATAAACTG CCAATGGAGC 7200

TACAAAATCT CCATTTTTAG TGTAGACGTA TGCTTCAGCT AAACTAATTG CGTTGGTTGC 7260

AATGAATTGT TTTTGATATT CCTTGACATC CAAATTTAAA ACATCAAAAT AATTTTCCAT 7320

TGTAACATCT CTTAGTTCAA TTGTCATAGT TTTGCTCCTT GTTAGAGGTT ATCATTGGCG 7380

CAAAATAATG TTTACGGCAA ACTGAGATAT AGGTTTCGTT ACCACCAATC TGGATCTGTT 7440

CTCCATCGTA AACGGGCAGT CCATCCTGTG TTCGCAACAC CATGGTCGCC TTTTTCTTGC 7500

AATACTGACA GATGGTCTTG ATTTCGTCAA TCTTGTCTGC TAAAAGCAAG AGATATTTGG 7560

AACCTTCGAA CAATTCATTG CGAAAGTCAT TTTTCAAGCC AAAAGCCATG ACGGGTATGT 7620

CTAACTCGTC CACAACACGA GCTAGGTCGT AAACATGGTG GCGTTTGAGA AACTGGGCTT 7680

CATCGACCAA AACACAGTAA GGTTTTTCTG GTAGGTCTCG GATATAGCCA AAGATATCCG 7740

TTGTTTCCTC AATCGCAAgG GCAGGGCGTT TCATGCCAAT TCGACTCGAC ACATAGCCAA 7800

CGCCGTCACG CGTATCCAGA GCCGAGGTCA TAATCACAAC ACCTTTTCCT TGCTCCTCGT 7860

AGTTATAGGC CACTTTGAGA ATCTCAATCG TTTTACCAGA GTTCATGGTC CCATAACGAT 7920

AGTACAACTG TGCCATGTTT CTTGCTTCAC GTCCATTTCT AAATTTTTGC TACATTCTAG 7980

TATATCATAA TTTTCTTAAG CTTTAAACGG CAAAATGTGG TAAAATAGAA GAAATCAAAA 8040

ACTAGTGGAG GAAGCTATTA TGCCATTTGT ACGCATCGAT TTATTTGAAG GACGCACGCT 8100

CGAGCAAAAG AAAGCTCTTG CTAAGGAAGT AACGGAAGCA GTTGTCCGCA ACACTGGAGC 8160

CCCTCAATCT GCTGTCCATG TCATCATCAA CGACATGCCA GAAGGAACTT ACTTCCCACA 8220

AGGGGAAATG CGTACTAAAT AAGCTAGCTT AAGCAGAATT GCTTAGGCTT TTTCAATCTC 8280

CAAGTAGCAT TCATTGAAGA AATATCCTAA ATTTGTTACA ATTTGAAAAG AAACTTGGAG 8340

AATTTCCAAG AAAAGAGCTA TTAATTAAAG GAAACATTAT GATTACACGT GAATTTGATA 8400

CCATCGCTGC TATCTCTACT CCACTAGGTG AAGGGGCTAT TGGTATTGTC CGCCTGAGCG 8460

gAACAGACAG TTTTGCTATT GCGCAAAAGA TTTTTAAAGG AAAAGACTTG AACAAGGTTG 8520

CCAGCCACAC TCTCAACTAC GGTCACATTA TTGATCCTCT GACTGGTAAA GTCATGGACG 8580

AGGTTATGGT TGGGGCTATG AAGTCTCCAA AGACCTTCAC TCGTGAGGAT ATTATCGAGA 8640

TTAACACCCA CGGTGGGATT GCGGTGACCA ATGAAATTCT CCAGCTAGCT ATTCGTGAAG 8700

GGGCTCGGTT GGCAGAACCT GGTGAATTTA CCAAACGTGC TTTTTTAAAC GGTCGCGTAG 8760

ACTTGACACA GGCAGAGGCT GTGATGGATA TCATCCGTGC CAAGACTGAC AAGGCCATGA 8820

ACATTGCGGT CAAACAATTA GACGGCTCCC TTTCTGACCT CATTAACAAT ACCCGTCAAG 8880

AAATCCTCAA TACACTTGCC CAAGTTGAGG TCAATATCGA CTATCCTGAG TATGACGATG 8940

TTGAGGAAGC CACTACTGCT GTTGTCCGAG AGAAGACAAT GGAGTTTGAG CAATTACTAA 9000

CCAAACTCCT TAGGACAGCA CGTCGTGGTA AAATCCTTCG TGAAGGAATT TCAACGGCTA 9060

TCATTGGACG TCCCAACGTT GGGAAATCAA GCCTTCTCAA CAACCTCTTG CGTGAGGACA 9120

AGGCTATCGT AACAGATATC GCTGGGACAA CACGAGATGT CATCGAAGAG TACGTCAACA 9180

TCAATGGTGT ACCTCTCAAA TTGATTGATA CAGCCGGTAT TCGTGAAACG GATGATATCG 9240

TTGAACAAAT TGGAGTTGAG CGTTCGAAAA AAGCTCTTAA GGAAGCTGAC CTAGTTCTGC 9300

TAGTACTAAA CGCTAGTGAA CCACTAACCG CCCAAGATCG CCAACTCCTA GAAATCAGTC 9360

AGGAGACTAA TCGCATTATT CTTCTTAACA AAACTGACCT GCCTGAAACG ATTGAAACTT 9420

CGGAACTACC TGAAGATGTC ATCCGCATTT CAGTTCTTAA AAATCAAAAC ATCGATAAAA 9480

TCGAAGAGAG AATCAACAAC CTCTTCTTTG AAAATGCTGG TTTGGTTGAG CAAGATGCTA 9540

CCTACTTGTC AAACGCCCGT CACATTTCCT TGATTGAGAA GGCCGTTGAA AGCCTACAAG 9600

CTGTTAACCA AGGTCTTGAA CTAGGGATGC CAGTTGACTT GCTTCAAGTT GACTTGACCC 9660

GTACTTGGGA AATTCTAGGA GAAATCACTG GAGATGCTGC TCCAGATGAA CTCATCACCC 9720

AACTCTTTAG CCAATTCTGT TTAGGAAAAT AAGAAAAATC CATGATCCTT CATTCGGTCA 9780

TGGATTTTAG GTTCTATAAT ATTTGTAGTG GGTAAATCCA CTATAGATAT TATGGAGCCT 9840

ATTTTATTGT AGAAAAAAAG TCCCATATGA CCTATAATGA AAAGCGACAA AACAACTCAT 9900

TAGAAAGAAT CATATGGAAC AATTACATTT TATCACAAAA TTACTAGACA TTAAAGACCC 9960

TAATATCCAG ATTTTAGACA TCATCAATAA GGATACACAC AAGGAAATCA TCGCCAAACT 10020

GGACTACGAC GCCCCATCTT GCCCTGAGTG CGGAAACCAA TTGAAGAAAT ATGACTTTCA 10080

AAAAACCTTC TAAAATTCCT TATCTTGAAA CGACTGGTAT GCCCACTAGA ATTCTCCTTA 10140

GAAAGCGTCG ATTCAAGTGC TATCACTGTT CAAAAATGAT GGTCGCTGAA ACTTCTATCG 10200

TCAAGAAGAA TCACCAAATC CCTCGTATCA TCAACCAAAA GATTGCTCAA AAGTTAATTG 10260

AAAAGATTTC TATGACTGAT ATTGCCCATC AGCTTTCCAT CTCAACTTCA ACTGTTATTC 10320

GTAAGCTCAA TGACTTTCAC TTTAAACATG ATTTTTCTTG TCTTCCTGAG ATTATGTCTT 10380

GGGATGAGTA TGCTTTTACA AAAGGGAAGA T 10411

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2393 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

GTTTTGGGTT CTGGAAATTA TCAGATGGTT GGAAAAGCCG TCCACATCAA GATAGTGTTC 60

GGAGATTTAA GTTTAAATTG AAGAAACTAA CACAGAGGAA ATGGAGTATA GACCTAACAA 120

GACGTATTGA GCAACTGAAT TTGTCTATTC GAGGATGGAT AAACTATTGC TCATTGGGAA 180

ATATGAAAAG TATAGTCGCC AGCATAGATG AGCGCTTGCG TACTCGCCTA CGAGTGATTA 240

TCTGGAAGCA ATGGAAGAAG AAATCGAGAC GATTATGGGG ATTGCTTAAG TTAGGAGTTC 300

CTAAATGGAT AGCAGATAAG GTATCTGGCT GGGGCGACCA TTATCAATTA GTAGCTCAGA 360

AGTCGGTACT TAAACGTGCT ATATCAAAAC CAGTCCTGGA AAAACGTGGA CTGGTTTCGT 420

GTTTGGATTA TTACCTTGAA CGACATGCGT TAAAAGTTAG TTGAACCGCC GTATGCCAAA 480

CGGCACGTAC GGTGGTGTGA GAGGGGCTAG AGATTATCCC CTACTCGATT AACTCCCCTG 540

AAATTTATTT TAATTATGCA AATTTCACGT ATTTTTGATG CTGAGACGAC GATCCTGGGA 600

ACTTTTCAGA TATTTTTTTG ACTATCTAAA TCTATCATTA GAAAAGCTTA GAGCGCCAAA 660

GGATTTGAGC GTTTTTCTGA TTTTTAAGAC TTTTTCCAGT CTCTTTTTCG ATTGAAGATG 720

TAATTATTCT ACTAACTAAC TAACTTCTTA GTACTAGCCA ACAACGATAA TCATAATTCC 780

TCCTAAAATT AGGAATAATA AAGGCAATAG TTTTTGTTTT TTCATGTAAA AAACCTCACT 840

TTTGTTTTCT GCTATTTTAT GCTAAAATAT TAAAAATCAA ATTTAATTCC AAAGTTTGTA 900

ACTAAAGGGG GAGCGCTACA TGTCTAATTC ATTTGTCAAG TTGTTAGTCT CTCAATTATT 960

TGCAAATTTA GCAGATATTT TCTTTAGAGT AACAATCATT GCTAACATAT ACATTATTTC 1020

AAAATCAGTA ATTGCCACAT CACTAGTTCC TATCTTAATA GGAATATCCT CTTTTGTTGC 1080

GAGTCTTTTA GTTCCGTTGG TTACTAAAAG GTTAGCGCTA AATAGGGTTT TATCTTTATC 1140

TCAATTTGGA AAGACTATAT TATTGGCGAT ACTGGTAGGA ATGTTTACCG TAATGCAATC 1200

CGTAGCGCCT TTGGTGACCT ATCTATTTGT TGTTGCAATT TCCATACTAG ATGGTTTTGC 1260

AGCACCCGTT TCCTATGCTA TTGTGCCACG CTATGCGACC GATTTGGGTA AGGCTAATTC 1320

AGCCTTATCA ATGACTGGTG AAGCTGTTCA ATTGATAGGT TGGGGATTAG GTGGACTCTT 1380

GTTTGCAACA ATTGGTCTGT TACCTACCAC GTGTATCAAT TTAGTCTTGT ATATCATTTC 1440

TAGCTTTCTG ATGTTATTTC TTCCTAACGC TGAAGTGGAG GTGTTAGAGT CAGAAACTAA 1500

TCTTGAAATT TTGCTCAAAG GTTGGAAGTT AGTTGCTAGA AATCCTAGAT TAAGACTTTT 1560

TGTATCAGCA AATTTATTGG AAATTTTTTC AAATACGATT TGGGTTTCTT CCATTATACT 1620

TGTTTTTGTA ACGGAGTTAT TAAATAAAAC GGAAAGTTAC TGGGGATATT CTAATACAGC 1680

ATACTCTATT GGTATTATAA TTAGTGGCTT AATTGCTTTT AGGCTATCTG AAAAGTTCCT 1740

TGCTGCTAAA TGGGAACCCC AATTATTCAC CCCAAATCTA AAAACCATCC AGAATCCTTG 1800

CCTTAGCTTA GATCCTGGAT GGTTTCTTTT TTCACCCAAT GGGTGTTTTT TACTAGACAA 1860

AAAAGAGTTT CCCCTTTATG GTATAAGTGT AGAAAAAAAC ACAAAAAGAA AGGAAACTCA 1920

CATGAACAGT TTACCAAATC ATCACTTCCA AAACAAGTCT TTTTACCAAC TATCTTTCGA 1980

TGGAGGTCAT TTAACCCAGT ATGGTGGTCT TATCTTTTTT CAGGAACTTT TTTCCCAGTT 2040

GAAACTAAAA GAGCGGATTT CTAAGTATTT AGTAACGAAT GACCAACGCC GCTACTGTCG 2100

TTATTCGGAT TCAGATATCC TTGTCCAGTT CCTCTTTCAA CTGTTAACAG GTTATGGAAC 2160

GGACTATGCT TGTAAAGAAT TGTCAGCTGA TGCCTACTTT CCAAAATTGT TGGAAGGAGG 2220

GCAGCTTGCT TCACAGCCAA CCTTATCCCG TTTTCTTTCC AGAACTGACG AGGAAACAGT 2280

CCATAGTTTG CGATGCCTCA ACCTTGAATT GGTCGAATTC TTTTTACAGT TTCACCAGCT 2340

AAACCAACTC ATTGTAGATA ACGATTCTAC CCATTTCACA ACTTATGGCA AGC 2393

(2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4762 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

TTTGTATCTT TTTAGGTCTC TTTCAATCCA AACCCTTTAA ACTATACGTC ATTTCGGTTC 60

CTGCAAGTCT TGTGGTAATT TTAGGTTTGA TTTTACTTTT CTTTTCACAA GAGCCTCTGC 120

ACGCTTCTTA TTTGATGGTC GTCTTCCCTG TTTTCCTACT TTTATTGGTA ACCAATATTA 180

AGAGTCAACA GAGGGGGCGT AGTGCTAGAA GAAGCCGAAG AGAAACGCCA TTATGCCTAT 240

GGAGTCGTTT CTTCAAAGGA AATCTATATC TGCTAGTTTT TGGGTTTGTC TATCTTTTGT 300

CTGTTCCTTT TTTGATGAAG TTTGTCCTTT ATCCAGTACC TTATCAAGAA CGTAATCGTC 360

TTGCTGATTT GGTAAAAGAG GAGACAAATA CGGAAGATGC TATCTCATGC ATGGGATGAT 420

ACTGCGACTC TTTATCGTAA GAGTGAGCGC TTGTCCCATC GGCGATTTTG TCCCCGTTGC 480

ACTATACAGC AACTGAGGAA AATCGTAATA AGTTACTTAA TGACTTGAAA GAAAAACAAC 540

CTAAGGTGAT TGTGGTAAAT GATAAGGTGG TAGTCTGGTC TGAAGTGGAA ACACTCTTAA 600

AAGAAAATTA CCAACAAGTA AAGACTGATT ACTCAGAGTT TAAAGTCTAT AAAATTAAAT 660

AACCAAATCA ATATCTTGTG TATTTTTAAA AATTTTAGGA TTTTTAACAC AAGATATTGA 720

TTTTTCTTTT TAGAGTGGTA TAATACTTTT TAGAAAGAAC ATTTTAGAAA AGAGCATGCA 780

TATGATTGCA CTAGAAGAAA AAATTACAAT TTTGCCAACT CTCTTCGTCG AGAAACGAGA 840

TGGGAGACGT GTTGTATTTG ATGTGGACAA GATTGACAAG GCTCTCCACA AGGCGGCTGA 900

CAAGGTTATG GATGTGACAC CCCTGGTTGA AAAATGCCTC AATGATCTGA CTGAGCGAAT 960

TATTACAGAA ATTCATAGTC GCTTTCCACA GGGAATTAAG ATTTACGAAA TTCAAAATAT 1020

CGTAGAACAT GAACTCCTTG AAGCCAAAGA ATATGCGCTG GCTGAGGAGT ATATTACTTA 1080

TCGGACACAG AGGGATTTTG AGCGCTCAAA AGCGACGGAT ATCAACTTTA GTATTCATAA 1140

ACTTCTCAAC AAAGACCAGA CAGTTGTCAA TGAAAACGCT AATAAAGACA GTGATGTCTT 1200

TAACACTCAG CGTGATTTGA CAGCAGGGAT TGTTGGGAAA TCAATCGGAC TGCAAATGCT 1260

TCCTAAGCAC GTAGCCAATG CCCACCAAAA GGGGGATATC CACTATCACG ATTTGGACTA 1320

CAGTCCCTAT ACCCCTATGA CCAACTGCTG TTTGATTGAT TTTAAGGGTA TGTTGGAAAA 1380

TGGTTTTAAG ATTGGAAATG CAGAGGTAGA GAGTCCCAAG TCTATCCAGA CTGCGACAGC 1440

ACAGATTTCT CAAATCATTG CCAACGTTGC TTCTAGCCAG TACGGTGGCT GTTCAGCTGA 1500

CCGTATCGAT GAAATTTTGG CGCCTTATGC AGAGAAGAAT TATCAAAAAC ATCTCAAAGA 1560

TGCAGAAGAG TGGGTATTGC CTGAAAAACA GGAAGATTAC GCTTGGAAGA AAGCGCAAAA 1620

GGACATCTAC GATGCCATGC AATCTCTTGA GTATGAAATC AATACTCTCT TCACTTCAAA 1680

TGGACAAACA CCTTTTACTT CGTTAGGTTT TGGTCTGGGA ACCAGTCGTT TTGAACGAGA 1740

AATTCAAAAA GCTATTTTAA ACATTCGCAT CAAGGGTCTT GGTTCAGAAC ACCGTACGGC 1800

TATCTTTCCT AAACTTATCT TTACGCTTAA AAGAGGCCTC AACTTAGAGG AAGGAACTCC 1860

CAACTATGAC ATCAAGCAGT TGGCTCTAGA GTGTGCAACC AAGCGGATGT ATCCAGACGT 1920

CTTGTCTTAT GATAAGATTG TTGATTTGAC AGGTTCTTTC AAGGTGCCTA TGGGCTGCCG 1980

TTCTTTCCTT CAAGGGTGGA AGGATGAAAA TGGTGTAGAA GTCAATTCAG GTCGCATGAA 2040

TCTGGGTGTT GTGACGGTTA ATCTGCCTCG TATTGCTCTT GAGTCTGAAG GTGATATGAA 2100

TAAGTTCTGG GAAATCTTCA ACGAGCGAAT GAATATCGCA GAAGATGCTC TTGTTTACCG 2160

TGTCGAACGC ACTAAAGAGG CGACACCAGC GAATGCTCCT ATTCTTTATC AGTACGGTGC 2220

TTTTGGCCAT CGTCTAGGTA AAGAAGAAAG TGTTGACCAG CTCTTTAAGA ATCGTCGTGC 2280

GACCGTTTCG CTGGGCTATA TCGGCTTGTA TGAAGTAGCG ACAGTTTTCT TTGGTAACAG 2340

CTGGGAAAGT AATCCAGATG CTAAGGAATT CACGCTAGAC ATCATTCACG ATATGAAACG 2400

CCGTGTAGAA GAGTGGTCAG ACCAATATGG CTACCATTTC TCTATCTACT CAACACCATC 2460

CGAAAGTCTG ACAGACCGTT TCTGCCGACT AGATATAGAC AAGTTTGGCT CTATTCCTGA 2520

TATCACAGAC AAGGAATACT ACACCAACTC TTTCCACTAC GATGTTCGTA AAAATCCAAC 2580

ACCGTTTGAA AAATTGGACT TTGAGAAAGT CTATCCGGAA GCAGGTGCGT CAGGTGGTTT 2640

CATCCATTAT TGTGAGTATC CAGTCCTTCA GCAAAATCCA AAGGCCTTGG AAGCTGTCTG 2700

GGATTATGCT TATGACCGTG TAGGCTATCT AGGCACCAAT ACTCCGATTG ACCGTTGCTA 2760

CAAGTGTGAC TTTGAAGGGG ATTTTGAACC AACTGAGAGA GGGTTTGCTT GTCCAAACTG 2820

TGGCAATAGC GACCCTAAAA CAGTAGATGT GGTGAAACGA ACTTGTGGCT ACCTAGGTAA 2880

TCCTCAAGCA AGACCGATGG TCAACGGGCG TCACAAGGAA ATCGCTGCGC GTGTCAAACA 2940

TATGAATGGT TCAACGATTA AAATAGCTGG GCATCAAGTA ACAAATTAGA AAGAAATGAA 3000

ATGGGAAAAT ATCAACTAGA CGATAAGGGG CGCGCACAAG TGACCCGTTA TCACGAGAAA 3060

CACTCTAAAG GTGGAGCTGG TAAGAAAGAA CGCTTGCTTA GCTTCAGAGA ACAATTTTTA 3120

AACAAGAACA AGAAAAAATA AAAGTGAGAG CCAGCTCTCG CTTTTCTCAT AGTGGGAGGT 3180

AAGGATGGAA TTACGCAGAC CAAGATTAGC GGATAAGAAA GCTGTTTTAG ATATGATGAC 3240

AGAGTTTGAA AAATTTCAGT CGCCTCACGA CGGCGGTTTC TGGGATACAG AGAACTTTGT 3300

GTATGAAGAC TGGTTAGAAA GCAATCAGGA ACAGGAAATG GGGATTAATC TGCCTGAAGG 3360

ATGGGTTTCT GCAATTCAGT TAGTGGCTTT TTCTGAGAAA GGTCAAGCAG TTGGATTTCT 3420

TAATCTCCGG TTGCGCCTCA GTAACTTTCT ACTAGAAGAA GGTGGCCACA TTGGCTACTC 3480

CATTCGTCCA TCTGAAAGAG GCAAGGGTTA TGCAAAAGAG ACTCTCCGTC AGGGCTTGCA 3540

AGTTGCTAAG GAAAAGAACA TCAAGAAAGC TCTGGTGACC TGTAGTGTGA ATAATCCTGC 3600

TAGCAGAGCA GTCATTCTAG CAAATGGTGG AATATTTGAG GATGCTCGCA ATGGAGTCGA 3660

GCGTTATTGG ATAGAGGTAG CGAATGAATA ATCCAAAACC ACAAGAATGG AAAAGCGAGG 3720

AACTTAGTCA AGGTCGTATC ATTGACTACA AGGCCTTTAA CTTTGTGGAC GGCGAAGGCG 3780

TGCGCAACTC TCTCTATGTA TCAGGCTGTA TGTTTCACTG CGAGGGATGT TATAATGTTG 3840

CGACTTGGTC TTTTAATGCT GGCATTCCCT ATACAGCAGA ATTAGAAGAG CAGATTATGG 3900

CAGACCTTGC CCAACCCTAT GTTCAAGGCT TGACTTTGCT GGGAGGGGAG CCTTTTCTCA 3960

ATACTGGGAT TCTCTTGCCA CTTGTTAAGC GGATTCGGAA GGAATTGCCA GACAAGGACA 4020

TCTGGTCCTG GACCGGCTAC ACTTGGGAAG AAATGATGTT GGAAACTCCA GATAAACTGG 4080

AATTCTTGTC ACTGATTGAC ATTCTTGTCG ATGGAAGATA TGATCGAACT AAGAGAAATC 4140

TTATGCTCCA GTTTCGAGGT TCATCTAACC AACGAATTAT CGATGTGCAA AAATCGCTCA 4200

AAAGTGGGCA AGTAGTGATT TGGGACAAGC TCAATGACGG AAAAGAAAGC TATGAACAGG 4260

TGAAGAGAGA ATGAAGAAAA AGGACTTAGT AGACCAACTA GTCTCAGAGA TCGAGACGGG 4320

GAAAGTCAGG ACACTGGGAA TATACGGTCA TGGAGCTTCA GGTAAATCAA CCTTTGCACA 4380

GGAATTGTAC CAAGCTTTAG ATTCTACTAC AGTAAATTTG CTAGAGACAG ATCCTTATAT 4440

CACCTCAGGA CGCCATCTGG TAGTACCCAA GGACGCGCCG AATCAAAAGG TGACAGCCAG 4500

TCTGCCAGTG GCGCATGAAC TGGAGAGTTT GCAGAGAGAT ATCCTTGCTT GCAGGCGGGT 4560

ATGGATGTCT TGACAATTGA AGAACCTTGG AAGGCTAGTG AGGTCTTGTC TGGAGCCAAA 4620

CCAATTTTGA TTGTCGAAGG GATGTCTGTT GGCTTTCTAC CCAAGGAACT CTTTGAAAAA 4680

ACCATCTGTT TCTACACGGA TGAGGAGACC GAATTAAAGC GACGCCTTGC TAGAGATACG 4740

ACTGTGAGAA ATCGCGATGC GG 4762

(2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:

CGTATCTTCT CGTTCTGAAA AAGTTGGTGA GTACATCTCT GAAGCAATGG AAAAAGTTGG 780

CAAAGACGGT GTCATCACCA TCGAAGAGTC ACGTGGTATG GAAACAGAGC TTGAAGTCGT 840

AGAAGGAATG CAGTTTGACC GTGGTTACCT TTCACAGTAC ATGGTGACAG ATAGCGAAAA 900

AATGGTGGCT GACCTTGAAA ATCCGTACAT TTTGATTACA GACAAGAAAA TTTCCAATAT 960

CCAAGAAATC TTGCCACTTT TGGAAAGCAT TCTCCAAAGC AATCGTCCAC TCTTGATTAT 1020

TGCGGATGAT GTGGATGGCG AGGCTCTTCC AACTCTTGTT TTGAACAAGA TTCGTGGAAC 1080

CTTCAACGTA GTAGCAGTCA AGGCACCTGG TTTTGGTGAC CGTCGCAAAG CCATGCTTGA 1140

AGATATCGCC ATCTTAACAG GCGGAACAGT TATCACAGAA GACCTTGGTC TTGAGTTGAA 1200

AGATGCGACA ATTGAAGCTC TTGGTCAAGC AGCGAGAGTG ACCGTGGACA AAGATAGCAC 1260

GGTTATTGTA GAAGGTGCAG GAAATCCTGA AGCGATTTCT CACCGTGTTG CGGTTATCAA 1320

GTCTCAAATC GAAACTACAA CTTCTGAATT TGACCGTGAA AAATTGCAAG AACGCTTGGC 1380

CAAATTGTCA GGTGGTGTAG CGGTTATTAA GGTTGGAGCC GCAACTGAAA CTGAGTTGAA 1440

AGAAATGAAA CTCCGCATTG AAGATGCCCT CAACGCTACT CGTGCAGCTG TTGAAGAAGG 1500

TATTGTTGCA GGTGGTGGAA CAGCTCTTGC CAATGTGATT CCAGCTGTTG CTACCTTGGA 1560

ATTGACAGGA GATGAAGCAA CAGGACGTAA TATTGTTCTC CGTGCTTTGG AAGAACCCGT 1620

TCGTCAAATT GCTCACAATG CAGGATTTGA AGGATCTATC GTTATCGATC GTTTGAAAAA 1680

TGCTGAGCTT GGTATAGGAT TTAACGCAGC AACTGGCGAG TGGGTTAACA TGATTGATCA 1740

AGGTATCATT GATCCAGTTA AAGTGAGTCG TTCAGCCCTA CAAAATGCAG CATCTGTAGC 1800

CAGCTTGATT TTGACAACAG AAGCAGTCGT AGCCAATAAA CCAGAACCAG TAGCCCCAGC 1860

TCCAGCAATG GATCCAAGCA TGATGGGCGG GATGATGTAA GCTTTCTATA GAAAACAACT 1920

TATAAAAAAC ACAAAAGGAG GGAATGACTA ACCCTTCTTT TTATAGGCTC TTTGTCAACT 1980

GTAGTGGGTT GAAGTCAGCT AAGCTCGAGA AAGGACAAAT TTCGTCCTTT CTTTTTTGAT 2040

GTTCAAAGCG ATAAAAATCC GTTTTTTGAA GTTTTCAAAG TTTCGAAAAC CAAAGGCATT 2100

GCGCTTGATA AGTTTGATGA GATTATTGGT CGCTTCCGGT TTGGCGTTAG AATAGTGTAG 2160

TTGAAGGGCG TTGATAATCT TTTCTTTATC TTTGAGGAAG GTTTTAAAGA CAGTCTGAAA 2220

AATAGGATGA ACTTGCTTAA GATTGTCCTC AATAAGTCCG AAAAATTTCT CCGGTTCCTT 2280

ATTCTGAAAG TGAAACAGCA AGAGTTGATA GAGCTGATAG TGATGTTTCA AGTCTTGTGA 2340

ATAGCTCAAA AGCTTGTCTA AAATCTCTTT ATTGGTTAAA TGCATACGAA AAGTAGGACG 2400

ATAAAATCGC TTATCACTCA GTTTACGGCT ATCCTGTTGT ATGAGCTTCC AGTAGCGCTT 2460

GATAGCCTTG TATTCATGGG ATTTTCGATC CAATTGGTTC ATAATTTGAA CACGCACACG 2520

ACTCATAGCA CGGCTAAGAT GTTGTACAAT GTGAAAGCGA TCCAACACGA TTTTAGCATT 2580

CGGGAGTGAA ACAGTCTGGG AGACTGTTTC AGCCTGAGCC TAGAAATTTG AAAGCGAAGC 2640

TGTTTAGCCA AGTCATAGTA AGGACTAAAC ATATCCATCG TAATGATTTT CACTTGACAA 2700

CGAACGGCTC TATCGTAGCG AAGAAAGTGA TTTCGGATGA CAGCTTGTGT TCTGCCTTCA 2760

AGAACAGTGA TAATATTAAG ATTATCAAAA TCTTGCGCAA TGAAACTCAT CTTTCCCTTA 2820

GTGAAGGCAT ACTCATCCCA AGACATAATC TTTGGAAGCC GAGAAAAATC ATGCTCAAAG 2880

TGAAAGTCAT TGAGCTTGCG AATGACAGTT GAAGTTGAAA TGGCCAGCTG ATGGGCAATA 2940

TCAGTCATAG AAATTTTTTC AATTAACTTT TGAGCAATCT TTTGGTTGAT GATACGAGGG 3000

ATTTGGTGAT TTTTCTTTAC CAGGGGAGTC TCAGCAACCA TCATTTTTGA ACAGTGATAG 3060

CACTTGAAAC GACGCTTTCT AAGGAGAATT CTAGAAGGCA TACCAGTCGT TTCAAGATAA 3120

GGAATTTTAG AAGGTTTTTG AAAGTCATAT TTCTTCAATT GGTTTCCGCA CTCAGGGCAA 3180

GATGGGGCGT CGTAGTCCAG TTTGGCGATG ATTTCCTTGT GTGTATCCTT ATTGATGATG 3240

TCTAAAATCT GGATATTAGG GTCTTTAATA TCGAGCAGTT TTGTGATAAA ATGTAATTGT 3300

TCCATATGAA TCTTTCTAAT GAGTTGTTTT GTCGCTTTTC ATTATAGGTC ATATGGGACT 3360

TTTTTTCTAC AACAAAATAG GCTCCATAAT ATCTATAAGG GATTTACCCA CTACAAATAT 3420

TATAGAGCCG AAAATTCACA TCTAATATAT GCAGACTACT TTGAAATGAA ATTAAAAAAA 3480

TTATTAAAGG ATGACACAAA AGTTTTTGAA AAATCTACAT TCAAATTTGT AGAAGGATAT 3540

AAAATATACC TGACAGAATC TAAAGAATCT GGAATTAAAC AAATGGACAA TGTCATAAAA 3600

TATTTTGAGT TTATTGAATC TAAAAGTATT GCTTTATATT TTCAAAAACG ATTAAATGAG 3660

CTGATAGATT AAATAGCATT TTCTCTGTTG AGATATTGTT TTTAAAATAT TGTACTAAAT 3720

GATTGATGCT ATGTGGAAAT ACAAAAAAAT GTTTTTGATA CGAAGTTGAC CTGTATTTTT 3780

TATACTAATC ATTTTCGTAT TTTTTGTATT AAACGATATA AGTTTGTTGT AAACTTACAA 3840

GGAATAAAGA CATTAAAAAA TAACAGTATA TCTATTTGTT TTATATATTT TACGAATTCT 3900

GCATAAATCT CTTTCTAGTA ATGTGTTGTA ACTCTGCTAT AATAGATTTA TTCCTTTTTG 3960

TGTTTACACA ATTTATTTTA TAGTACCAAA AAAGGTCAGG ATTTTGTTCC TGACCTTTGA 4020

CAACTTTACC GATTCTTTAG TTCTACATAG CGCTTGTACC AAATGTTTAC ATAGGCTTCT 4080

GAGAAAGGAC CACGTCCATT GTTAATCCAA TCAACAAGAA TTTTGACATG TTCTTTTAAA 4140

ATATAGTCCA AGTCATCAGA ATAATTCATT TTGCGTTTGT GACGCTCGTA CTCTTCAACG 4200

TCCAAGAGAC GTTTTTCCCC ATCTGTAAAA ATTTTAACAT CCAAATCGTA ATCAATATAC 4260

TTCAGTGCTT CTTCATCCAG ATAGTAGGGG CTAGCCATAT TGCAATAGTA AGAAGTTCCA 4320

TTATCACGAA TCATGGCAAT GATATTAAAC CAATATTTCT TGTGAAAGTA AACAATAGCC 4380

GGTTCTCGAG TGACCCAACG ACGACCATCA CTTTCGGTAA CAAGTGTATG ATCGTTGACA 4440

CCAATAATGG CGTTTTCTGT TGTTTTTAGT ACCATGGTGT CCCGCCAAGT TCGGTGGAGA 4500

CTCCCATCAT GCTTATAACT TTGAATTGTA ATAAAGTCGC CTTCTTTTGG AAGCTTCATA 4560

ACTAACCAAC TTTCTACAAT TTATAAGTTT ATCATTTACT ATTGTACCAT AAAATTACCC 4620

AAAATCTGTG AATTTCACTT GGAAATATTA AAGATATTCT CTAAGAGCGC TTGCTATATC 4680

CGAAAAATCG TAGCCCTTTC GTGCTAAAAC TTGAGTTAAA CGCTGCTTCA GTTCGTATCC 4740

TTCATACTTT CGGGCATACT TAGTATATTG CTTATCAAGT TCCTTGAAGA TGAGTTCCTG 4800

AGTCGTTTCT TCATCAACTT GACTATCCAA TTCGTCAAAG GCAATTTTAG CATCAAAATA 4860

AGAGAAGCCC TTGTTAGTCA AGTTCTGGAT AATCTTATCT TGCAGGGCAC GAGCTGGAAG 4920

TTTTCCCTCA TATTTTTTCA ATAGTTTATT GGCTACACGT TGAGCAACTT CCGAAAAATC 4980

AAAATCATTC AAGATTTCTT CTATAGTAGA TTTTGAAATT CCTTTTTGTG CTAATTTCTG 5040

AGTCAGTACA TAAGGTCCCT TGTCTCCTGA AAGTTGATTG GCATTGATGA TAGCATAAGC 5100

GTACTGGCTA TCATTAATCC ACTTCTCTTC TTTAAGATTA GCAATGACTT GAGAAACGAT 5160

GTTTTCATTA ATATCATATT TTTTCAGATA TTCTCTGACC TCTTTTTCAG TACGTGCTTT 5220

AAAGGATAAG TGGTAGAGGG CCAGATTCTT ACCATAAGAA AATTGAGCAA AGTCTTGAAT 5280

CTCTTTCAAT TCCTCTTCGC TTATCACCTT ATCTCTCGAT AACATAAAAC GAACAATTGT 5340

GTCTTCGGTG ATATAGCATT TGTCG 5365

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3636 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

TTTCCAGAAA GAAGTTGAGT AAAGTCTTTA TCAAAGAGAA TGACTTCCGT ATTGGAACTG 60

ACATTAGGTT TTATTTCTAC TTTACTAGCG TCCGCCCTAG CATTTTCTAA ATCTTTAATC 120

TCTTCTGTTG CCCTATTTAT AGCCAGCTGA ATAACTGCTT GAGGATTTTC ACTCAGTCCA 180

TGAAGCTTAT CGTCCACCGA AGTATAAAGA CTCGAATGCA TGACTTGTAA AATAATCAGA 240

GTCATTGTAG AAAAAATCAG GGTGAAGACA CCGAAGTTGC GGATAAAATA ACTAAAGTCA 300

TCCGCATACC ATGTTTTTTT AAGTTTACTG AACATCTTTT AAAAGATACC CAACACTACG 360

CAAAGTTTGC AAATTCTCTG CAAAAGTGGT TCCCTTTAAT TTCTTACGGA CTTTTGAAAC 420

ATAGACTTCG ACAACCGAAA TCGTTGTATC ACTATCAAAT CCCCATAGAC GGTCAAAAAT 480

CTGCGTCTTA GGCAAAATCA CATTTTGATT TTGAAGGAAA TAAACTAGTA AATCGAACTC 540

TTTCCCCAGC AATTCGACAG GAGTATCTTC AACTTTAACG GTATTGGTTG ATAAATTAAC 600

CACGATATTC CCATAAGTCA AGGTGTTTTC ATTAAACTTC CCTGAACGTT TGAGAAGGGC 660

CTGAATCCGC ATTTTAAGTT CTTCTAGGTA GAAAGGTTTG GTCAGATAAT CATCCGCTCC 720

CAGTTCAAAT CCATGTCCCT TGTCATCCAA ACTTTCCTTG GCAGTCATAA TCAGAACTCG 780

TGTCGTAATT CCCTTTTCAC GCAATTCTTT TAAGACTTGG AAACCATTTT TTTCTGGCAA 840

CATCAAATCC AGCAAAATCA AGTCATAGAC ACCACTCTCA GCTTCGTAGA GACCTTCTTC 900

TCCATCAAAT ACCTGCATAA CATCCGCAAA ATCGTCTAAA AAGTCAAATA CTGAATTTGA 960

CAGACCTAGG TCATCCTCAA CCAATAAGAT TTTTATCATG AGAAACTCCT CCTTATTAAA 1020

ACTATTATAC CAAATTTGCC TTAAAAAAAA CTCAACTCTC TGCATTTTAC ATGAGATAGC 1080

TGAGTTTTCT TTTTATTTTA GGCTTATTTA TGCATTTCCG TATTGAAGAA CAACTGCTTC 1140

GACTGCAGCT TTTTCACGGC TAATCAAGTC AACACGCGCT GCAATTTCCT TGATTCCCAT 1200

ACCGATGTTA CGGCTAAGAG CAAGGTCAGA AAGTTGCGGT TCAAAGAACT CCTTGTATTC 1260

CGCCAAGCGT TGCTGAGTCT TAAATACATG AGCAGGAAGG ATAACAAAGC TATCAAAGCT 1320

CATATCTCCT CCAAGGGCTG CCTTAATCCA AGCCCAGTTT TCACGCGCCC AAGACCAAGC 1380

TGTTTTCTGA GTTGCTTGAT GAGCTAGGAA TTGGTAATAC CAAGCAGACA AGTCCTGTGG 1440

TTTGACCACA AATTTGTCCT TCCAAGAAGT AATCAGGTTT TGGATATTAT CCGCATCTGT 1500

ACTGTATGCA AGAGCTGCTG CCAACTGGCG TTTAAAGACA GCATCTGTTG CGTGAGTATA 1560

AGTATCAAGA TAAAGTGCTA ACAAGTCTTT AGTCTCATGA TGTTTCATCT CATTAATCAG 1620

AACTTGTGAG CGAATAGCTG CTGGGAGTCC TGCAAGATTC TCCTTGTGTG TTGCGAAGAT 1680

TTGGCTAGCG ACTTGACTAG CTTCTGCATC ATTTGAGCGA ATCATCATCG AAACAGCCAG 1740

CTGACGAACC AATTCATCCT CATCTGATTC TCCGTCTTTA GCTTCAAAAC CAAGACGGTC 1800

ATAGTTATGA CGAGCCAATT TAGCAACCAG TCCTTTGAAG GCTGTTTCAG CATCCGTTCC 1860

TTCATCAATA AAGCGCTCAA GGGCTGAAAT CACTTGAGAA ACAGCTGAAA CCACCAGATA 1920

AGACTCTTCC TTAGCAAGTT TATCAAGAAC TGGAAGCAAG TCTGCATAAG AAATGTGCCC 1980

TGCCTCAGCC AACAAACGAC GTTCTTGAAC AATTTGCAGT TTGCTTGTGT TATCAAGTGT 2040

CTCTAGCTCA GCAAGAACAG CTGCTAACAA GTCTCCTTGA TAGTCGGTAA TATAGTGGGC 2100

AGTATTTTCA GTGTTGAGAC GAAGAGCTCC TTCATTTTCA GCAAGAAGAG CTGCGTAGCC 2160

AGGGATTTCG ATACTTTCAG TTTCGAGTGT ATCAGGCAAG CCTTTCCAGT TGCTATTGAG 2220

GGGCACCACC CAGAGACGGT TCTTGTCTTC GTTCTCACCG ATGAAGAATT GTTTTTGTGA 2280

AATCTTCAAG ACATCATTTT CAACTTTAAC AGTAAGAACT GGGTAACCAG GCTGTTCCAA 2340

CCAAGAATCC ATGAAGGCTG CGACATCACG TCCTGACGCT TGACCAAGGG CATCCCAAAG 2400

GTCACTACCA ATGGTGTTGC TGTATTGGTG TTTTTCAAAG TAGGCGTGCA AACCTTTAGC 2460

AAAATCAGCA TCTCCTAGCC AACGGCGAAG CATGTGCATG AGACGGCTTC CTTTGGCATA 2520

GACGATAGCG CCGTCAAAGA GTGTATTGAT TTCATCTGGA TGTTTAACTT CGACGTGGAC 2580

AGACTGAACG CCATCAGTAG CGTCACGTTC AAGAGCAAGA GGTACTCCAC CTGTTTGGAA 2640

ATCTTCAAAG ATATTCCAGC TTGGTTCGAT GGTATCCACA CAGACGTATT CCATCATATT 2700

AGCGAAACTT TCATTGAGCC AAAGGTCATC CCACCATTTC ATAGTCACGA GGTTCCCAAA 2760

CCATTGGTGA GCCAATTCAT GGGCCACAAC AAGGGCAACT TGTTGACGGC TAGCAAATGT 2820

AGAGTTCTCA TCGACAACCA AGTAAACTTC ACGGTAGGTC ACAAGACCCC AGTTTTCCAT 2880

AGCACCAGCT GAGAAGTCAG GAAGGGCGAT GTGGAGAGAT TGAGGAATTG GGTACTTAAC 2940

TCCATAGTAA TCTTCGTAAA ACTCGATAGA GCGAACAGCG ATATCCAGTG AGAAATCAAG 3000

ATTTGAAAGT GGATGTGCTT TGGTTGAGTA GACACCTACC AGGGTACCAT TTTTAGTTTT 3060

AGCGGTCACC CCTTGCAAAT CACCAGCAAC AAAGGCCAAC AAGTAAGAAG ACATGCGAGG 3120

TGTTGTCTCA AACTTCCAGA TACCTGTTTC CTTACGGTTT TCAACATCGA TTTCTGGCAT 3180

GTTTGACAAG GCCAATTCAC CTTCTGCTTG GTCAAAGCGA AGAGAGAGGT CAAAAGTTGC 3240

TTTGGCTTCA GGCTCATCCA CACATGGGAA AGcTTCGCGC GCAAAATGGC TCTCGAACTG 3300

AGTAGACAAG ACCTCCTTCT TGACTCCATC AACTGTATAA TAAGAAGGGT AAATCCCTGT 3360

CATGTTGTCT GTAATTTTAC CAGAAAAGGC AAGAACCAAT TCAACTTGAC CAGCCTCAGC 3420

CAATTCGATA TGAAGGGCTT CATTGTCATG GTCAACTGTA AATGGACGAG CTTGACCTGC 3480

AACTTCTACA GAGGTGATTT CCAAATCTTT TTGGTGGAGG GAGATGCCGT CACTCTGTGC 3540

TTGACCAGTG ATGGTCACTT TCCCAGAAAA AGTCTTGGTC TCACGACTCA AATCTAAAAA 3600

TAAATCATAA TGTTCAGGAA CAAATTGCTT AATGGG 3636

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5066 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

ATAGCGTGTA ATAATCGATT TTAGAGGTAC CATAAGCCAC CTCCTACAAA TAGAAACCGA 60

TATAAATCAA TGCCTTCCAC CCTTAGACTT CCCTAGTTCC TGTCTCAAGC GAAACATTTC 120

TTTGAAACAG GAATAAGTTA ACCAATTCAT ACCAATAGCT AGCAGAATAA AAAGAAACCA 180

AATGCCCCAT AACTTGATAT CTGTCACATT TCTCAAGACG GTATTGAAAA ACAGAACTGA 240

AACAACTGTC CAAGCAAGGC TAAAAAGAGA ATAGAAGGGG ATGTAAAACC AGTAAAAATA 300

ATAAAAAATT GGAAAAAACT TACTATTTCT GTTGGCCTTT TCAATCCAGT TATCAAAATA 360

AAAGTACGGT GCTAAAAGTA AGAATTTAAA CAAATGTTCC ATCACCGACA TCCCCCCTTC 420

TTTTGATAGC GTTTTCTATT ATTTTATTAT ATCAAAAAAA TCCGGAACTG TCATTCCAGA 480

TTCTACTTTT TTATTTGCGT TTTCTTGCGA TGAGATGAAT CGGTGTTCCC TCAAAAACAA 540

AGGCCTTGCG GATTTGATTT TCCAAGAAAC GCAGGTAAGA AAAGTGCATG AGTTCTTCTT 600

CATTGACAAA GATGACAAAG GTTGGTGGTT TGGTTGCCAC TTGGGTCGCA TAGAAAATCT 660

TGAGACGTTT TCCTTTGTCT GTCGGTGTTG GGTTGATGGC AATGGCATCC ATGATGACAT 720

CGTTCAAGAC AGCTGATGGA ATACGTGTAT TTTGACTTTC GCTGATTTGC TTAATCATCT 780

CAGGAAGTTT GTGGAGACGT TGCTTGGTTA AAGCTGATAC AAAGATAATC GGTGCGTAAG 840

GCAGGTATTG GAACTGCTCA CGGATATCTT CTTCCCAGTT TTTCATAGTG TGGTTATCTT 900

TTTCAAGCGT ATCCCACTTG TTGACCACGA TAATCATCCC TTTACCAGCT TCATGGGCAA 960

ATCCTGCGAT ACGCTTGTCG TACTCACGAA TGCCTTCTTC CGCATTGATG ACCATCAAGA 1020

CCACATCTGA ACGGTCAATA GCACGCATGG CACGCATAAC AGAGTATTTC TCAGTATTTT 1080

CATAAACCTT ACCAGACTTA CGCATACCAG CCGTATCAAT CATGGTAAAC TCTTGACCAT 1140

CTGTATCTGT AAAGTGGGTA TCAATGGCAT CACGAGTTGT TCCAGCAACA GGACTAGCAA 1200

TAACACGGTC TTCTCCCAAG ATAGCATTGA TCAAGCTTGA TTTTCCAACG TTAGGACGAC 1260

CAATCAAGCT AAACTTAATG ACATCTGGAT TTTCTTCCTC ATATTCATTT GGAAGATTTT 1320

CTACGATCGC ATCTAGCACA TCCCCTGTAC CGATTCCATG GACAGATGAG ATAGGCAATG 1380

GTTCACCCAA ACCGAGAGCA TAGAAATCAT ATATATCATT TCTCATCTCA GGGTTGTCCA 1440

CCTTGTTGAC TGCGAGGATA ACTGGTTTGT GGGTCTTATA AAGCTTACGA GCTACGTATT 1500

CGTCTGCATC AGTAATTCCT TCCTTACCAG ACACGACAAA AACGATAACA TCTGCTTCTT 1560

CCATGGCAAT TTCTGCCTGG TGCTTGATTT GTTCCATGAA AGGAGCATCG ACATCATCAA 1620

TTCCTCCTGT ATCAATCATG CTAAAAGAAC GATTGAGCCA CTCACCCGTT GCATAAATAC 1680

GGTCACGTGT CACTCCTTCG ACATCTTCTA CAATGGAGAT TCGCTCACCA GCGATCCGAT 1740

TAAATAGGGT TGATTTCCCA ACATTGGGAC GTCCTACAAT GGCAATAGTT GGTAGGGCCA 1800

TAATTTCTCA CTTTCTACAA TAATTTCTTC TGTTCAAGAT TTTTTCTAGT TGAGCTTGGT 1860

TCAGCTTGAC CAAACTGTTC TGCTAGGCGC TGACTCCAGC TTGTGGTCGC ACGCGCCCCA 1920

GCATAGTCAG CCTGAACACG GTCATAAGCT TGGATTGCCT CAGTTGACTG TTCTTGGTAT 1980

TCTTCCTCAA AGACAACATT CTCTAGTGGC AGTCTCGGTT TCATATCATG ATGTTGATTT 2040

GGCACACCCA GTGCCATCCC AAAGACAGAA TAGGTGTAGT CAGGTAGGTT AAAGAGCTCT 2100

GCCACTTCTT CAGACTTGTA TCGAACCAAA CCGATAATCA CACCACCATA GCCCAAGCTT 2160

TCAGCTGCCA ACAAGGCGTT TTGTCCAGCA AGAGCTGCAT CGACCGAACT AATCAAGAGA 2220

CCTTCCACAC CTTGGGGTTG GAAGGTGTCG GTATGAAGTC GGGCTCCCTT TTCTGCTCGG 2280

TTCAAATCTC CGACAAAGAG AAGGAAAACA GCAGACTGGC GAATGGCTTC TTGAGGTACC 2340

AATTCATACA AGGCATCTTT CTTCTCTTGA CTTCGTACCA CAATCACAGA GTAGGATTGG 2400

AAATTCTTCC AAGATGATGC CATCTGGGCT GCTGTCAAAA TCTCATTTAA GTCTACTTGG 2460

GGAATTTCTT GCTCTTTAAA CCTGCGCACT GAAGTATGAG CCTTCATCAA TTTAATGGTT 2520

TCTGTCATCG ACGGTTTACT CCTTCTAAAC GAGTCTCCTC AGCCAAATAA CGGATGCGTT 2580

CCATGACCCG TCTGGCTTCC CAGGTTTCGT CATTTCCATG TTTCACTTTC GCAAAATGCT 2640

TCTCCAAATC TTCAAAGTTG AAGTTGGATG TGAAAAAGGT CGGTAAATTT TCCTGCATCC 2700

GATATTGGAG AATGACCTGC AGGATTTCGT CACGCACCCA AACGGTTGAT TGCTCGGCGC 2760

CAATATCATC TAAAATCAGG ACCTCAGACA GCTTAATCTC ATCCACCAAG GTCTTAACAT 2820

TGCCATCACT GATAGCATTT TTGACATCAA TGACAAAGCT AGGATAGTGG AGGAGAGTTG 2880

ATGAAACACC ACGTTTTTCT GATAAATCAT GAGCTAAGGC CGCCACCATG AAACTTTTAC 2940

CCACACCAAA GTCTCCATAT AAGTAAAGAC CTTTTCGAAT AGCTGGATAT TGCTCCACGA 3000

AGGCTAGTAG CTTTTCAAAA ACTGGTAAGC GCCCCAAATC ATCCAAGTCA ACTTGAGCCA 3060

AACTAGCTTT CTTGAGACTG GCTGGTAGAT TGATTAACTT GAGACGGTTC TTAATAGCCG 3120

CTTCTTTTTC AGCCGCGATT AGCTCAGGAG TTTCTTCATA TGAAACATCT GCATAACCAT 3180

GATTCTTAAC CAAAATCGGC TTGTAGCCTT TGGCAATATA ATCCGTATCC CCACGGAGAA 3240

ACTTGTCACG CTCGGTGATG TACTGATTAA ACTTGGAGAT ACTGCGATTT AATTCCTTTG 3300

GAGTTAAGGA TTCTTGCTGG ATAAAGGCCG CAACATCAGG GTCCTTCATG ATTTTCTGGA 3360

CCAAATCTTG ATAATAAAAA CGGCTGGGTT GACGTTTGAG TACGTCTCCG ACACTTTCCA 3420

TCTAATCTCC TCCTTTTTCT AATCGAGCTA ATAGTTCTTG CTTCTTACGT TCTAGTTCCA 3480

GACGAGTTTC CTCGCTGGTT TCATTCTTAT ATTCAGGATT ACTCCATTTA GGAACATTGG 3540

TTTTTTCTGG GGCAGTCTGA TTCTGTTTTT GTGTTTTTGC TTTCTGCCCT CGATCACGAA 3600

TTCGTAAAAC GGCCTCTTCT GCCGAATGAA TCTTTTGATA GGCATAGTCA TTGGCTACCT 3660

TCATGGCATA TTTCTCATTG ATATTTGCCG AATCCACCTT ATTAAAGGTC AATAAGAGAA 3720

TAATATTGAT GACTTCGTCC AGTAAGCCCA AGCCAGCCAT CTGTTGCAAG AGTTCTCTTT 3780

CTGTTTGGGT AATGGTTCCC TTGCGTGTTT GCTTGATTTC TGCTAAGAAC TGCAGGGCAG 3840

TTTTACTTTT AGCTTCTTTG ATAATGGTCG CTTCCTTAAG ACTAAAGTCA GAGGAAACTG 3900

GTTTTTGAGC AATTTTTTCA CGCATGCGTT TGGTTGAAAT AACCTGGGAA ACAGCTGTTG 3960

ACTTGGCCAA TTGATAGGTT TCAAACCAAG TCCATTTCTT CTCCTCGGCA ATAGCAAAGA 4020

GGTTTAAGAC ATCGGACTGC TCATCCGCAA AACGAAGTCC ATCTCGAGCC ATCAGCTGGC 4080

GAAAATGTTC CAAGTCAAAA TCATTGGCCA CTTTCTTCTT GAGACCAAGG TCTTCTTGAC 4140

TGCCTAGTTC TGCCAATTCT GGAAAGACTT GATTGAGTGA GACAGGTATT TCTTCACCAT 4200

CAGCACTTTC AACTTTCAAA TCCTCCACAG CTACATCGCC AATCTTTTTC TCTAAGAGTC 4260

TGCGATAAAC AGGATGCCCC AAGAAGTCTT GACTAGATAG AGGAGCATGG AGGGCTAGCT 4320

GATAAACATC ACCCTTTTGA TAGAGGGTCA AGAGATTAAA AGCAGATAAG ATTTTCAATG 4380

ATTTTATCAG TCTATCCATC CCAAAGTTGA GATGGTTGAG AATGCTTGAA AAAAGATATT 4440

CCTTTCTACC ATTATCCCAA AAACTGATTG TATAAAGATA AAGGCTCAGT GCCTCCTGAC 4500

CGATAATCGG GAGGTAGCAC TGTACCAGAG ATGAGGTATC TTGCGACACC CGATTATTCT 4560

TTAGATAAGA AAAACGGTCA ATTGGCTTCA TTTATCTTTC CTTTTTCTTT TTAGAGGACT 4620

GGGTGATTTG TTGGAGCAAG CTCTCTAACT CACTGACATC CTTAAAACTA CGATAGACAC 4680

TAGCAAAACG TACATAGGTA ATCTCGTCCA ATTCAGCCAA CTCCTCCATG ACGAGTGAAC 4740

CAATGTCCTC ACTTTGAATT TCATTTTCAT TTCGACCACG GAGTTTCTGT TCGATACGAT 4800

TGACTACCAT GTTGATTTCA TCACTTGACA CAGGACGTTT CTGGGCTGAG CGGATAATCC 4860

CATTAAAGAT TTTATCTCTG GAGAATTGTT CCCGTGTGCC ATCTTTTTTA ACAACCACTA 4920

AGGTTCTTTC TTCTACTCGT TCGTAGGTTG TAAAACGGTG TTGGCATTCG TCGCACTCAC 4980

GTCTTCTACG AATGGTGTTC CCTTCTTCTG CTTGGCGACT ATCGATAACA CTTGACTTGG 5040

TAGCCCCACA TTTTGGACAG GGTACC 5066

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9607 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

CACTTGAAGT ATTTGAAACA GCTATGGAAA ACATCATGCC TGTACTTGAA GTACGTGCAC 60

GTCGTGTTGG TGGTTCTAAC TACCAAGTCC CAGTTGAAGT TCGTCCAGAA CGTCGTACAA 120

CACTTGGACT TCGTTGGTTG GTAACAATCG CTCGTCTTCG TGGTGAACAC ACAATGCAAG 180

ACCGTCTTGC AAAAGAAATC TTGGATGCTG CTAACAACAC TGGTGCAGCA GTTAAGAAAC 240

GTGAAGATAC TCACCGTATG GCTGAAGCTA ACCCTGCATT CGCACACTTC CGTTGGTAAG 300

ATAGGATGCG AAAGCGTTAA GAAAGTCCCA GAGAAAATAG GGAATCGAAG CAGGTTGCGG 360

TTGCAACCAA TGAGATTCAT CTTTTTCTCC AGACTTTTAG CTTGAGCTCA ACTAAATCAT 420

GATGCTAGGA ACGGTAAGGA TGCAAGGTAA AAATAGGAAA CTGACGCAGT ATTCGACGAA 480

TACAAGGAGT TTTATCTTTT TCACGCAGCA TCCCGTTCCA GCTCACATCG GCTAACTAAC 540

TTTAGCCCGG GTTCAAATTA GCTAAATCGA TTAGTATTAG CTATAACTCA GCTTACCATC 600

TCGTAAGTTG AAACCAACAA TAGCATGAAA ACATTGAGAA CGGGTAGGTC CTGCCTATCC 660

GTTTTTATTA AAATCGTGTT ATAATAGAAT AGAAATCAAA AATAAATAGG AGAAACAAAC 720

CTCATGGCAC GCGAATTTTC ACTTGAAAAA ACTCGTAATA TCGGTATCAT GGCTCACGTC 780

GATGCCGGTA AAACAACAAC TACTGAGCGT ATTCTTTACT ACACTGGTAA AATCCACAAA 840

ATCGGTGAAA CTCACGAAGG TGCGTCACAA ATGGACTGGA TGGAGCAAGA GCAAGAACGT 900

GGTATCACGA TCACATCTGC TGCGACGACA GCTCAATGGA ACAACCACCG CGTAAACATC 960

ATCGACACAC CAGGACACGT GGACTTCACA ATCGAAGTAC AACGTTCTCT TCGTGTATTG 1020

GATGGTGCGG TTACCGTTCT TGACTCACAA TCAGGTGTTG AGCCTCAAAC TGAAACAGTT 1080

TGGCGTCAAG CAACTGAGTA CGGAGTTCCA CGTATCGTAT TTGCCAACAA AATGGACAAA 1140

ATCGGTGCTG ACTTCCTTTA CTCTGTAAGC ACACTTCACG ATCGTCTTCA AGCAAATGCA 1200

CACCCAATCC AATTGCCAAT CGGTTCTGAA GATGACTTCC GTGGTATCAT TGACTTGATC 1260

AAGATGAAAG CTGAAATCTA TACTAACGAC CTTGGTACGG ATATCCTTGA AGAAGACATC 1320

CCAGCTGAAT ACCTTGACCA AGCTCAAGAA TACCGTGAAA AATTGATTGA AGCAGTTGCT 1380

GAAACTGACG AAGAATTGAT GATGAAATAC CTCGAAGGTG AAGAAATCAC TAACGAAGAA 1440

TTGAAAGCTG GTATCCGTAA AGCGACTATC AACGTTGAAT TCTTCCCAGT ATTGTGTGGT 1500

TCAGCCTTCA AAAACAAAGG TGTTCAATTG ATGCTTGATG CGGTTATCGA CTACCTTCCA 1560

AGCCCACTTG ACATCCCAGC AATCAAAGGT ATTAACCCAG ATACAGACGC TGAAGAAATT 1620

CGTCCAGCAT CTGACGAAGA GCCATTTGCA GCTCTTGCCT TCAAGATCAT GACTGACCCA 1680

TTCGTAGGTC GTTTGACATT CTTCCGTGTT TACTCAGGTG TTCTTCAATC AGGTTCATAC 1740

GTATTGAATA CTTCTAAAGG TAAACGTGAA CGTATCGGAC GTATCCTTCA AATGCACGCT 1800

AACAGCCGTC AAGAAATCGA CACTGTTTAC TCAGGTGATA TCGCTGCTGC CGTTGGTTTG 1860

AAAGATACTA CAACTGGTGA CTCATTGACA GATGAAAAAG CTAAAATCAT CCTTGAGTCA 1920

ATCAACGTTC CAGAACCAGT TATCCAATTG ATGGTTGAGC CAAAATCTAA AGCTGACCAA 1980

GACAAGATGG GTATCGCCCT TCAAAAATTG GCTGAAGAAG ATCCAACATT CCGCGTTGAA 2040

ACAAACGTTG AAACTGGTGA AACAGTTATC TCAGGTATGG GTGAACTTCA CCTTGACGTC 2100

CTTGTTGATC GTATGCGTCG TGAGTTCAAA GTTGAAGCGA ACGTAGGTGC TCCTCAAGTA 2160

TCTTACCGTG AAACATTCCG CGCTTCTACT CAAGCACGTG GATTCTTCAA ACGTCAGTCT 2220

GGTGGTAAAG GTCAATTCGG TGATGTATGG ATTGAATTTA CTCCAAACGA AGAAGGTAAA 2280

GGATTCGAAT TCGAAAACGC AATCGTCGGT GGTGTGGTTC CTCGTGAATT TATCCCAGCG 2340

GTTGAAAAAG GTTTGGTAGA ATCTATGGCT AACGGTGTTC TTGCAGGTTA CCCAATGGTT 2400

GACGTTAAAG CTAAGCTTTA TGATGGTTCA TATCACGATG TCGACTCATC TGAAACTGCC 2460

TTCAAGATTG CGGCTTCACT TTCCCTTAAA GAAGCTGCTA AATCAGCACA ACCAGCTATC 2520

CTTGAACCAA TGATGCTTGT AACAATCACT GTTCCAGAAG AAAACCTTGG TGATGTTATG 2580

GGTCACGTAA CTGCTCGTCG TGGACGTGTA GATGGTATGG AAGCACACGG TAACAGCCAA 2640

ATCGTTCGTG CTTACGTTCC ACTTGCTGAA ATGTTCGGTT ACGCAACAGT TCTTCGTTCT 2700

GCATCTCAAG GACGTGGTAC ATTCATGATG GTATTTGACC ACTACGAAGA TGTACCTAAG 2760

TCAGTACAAG AAGAAATTAT TAAGAAAAAT AAAGGTGAAG ACTAATCCGT CCTCACTCTA 2820

GAAGGAAGTC ACTTAGTGGC TTCCTTTTGT CTTTAGAAAA TACCTCTAAA TATGGTAAAA 2880

TAGTAGAAGA ATAATGTGAG GAAAATGAAT GTCAAATAGT TTTGAAATTT TGATGAATCA 2940

ATTGGGGATG CCTGCTGAAA TGAGACAGGC TCCTGCTTTA GCACAGGCCA ATATTGAGCG 3000

AGTTGTGGTT CATAAAATTA GTAAGGTATG GGAGTTTCAT TTCGTATTTT CTAATATTTT 3060

ACCGATTGAA ATCTTTTTAG AATTAAAGAA AGGTTTGAGC GAAGAATTTT CTAAGACAGG 3120

CAATAAAGCT GTTTTTGAAA TTAAGGCTCG GTCTCAAGAA TTTTCAAATC AGCTCTTGCA 3180

GTCCTACTAT AGGGAGGCTT TCTCTGAAGG TCCATGTGCT AGTCAAGGTT TTAAGTCCCT 3240

TTATCAAAAT TTGCAAGTTC GTGCTGAGGG TAATCAGCTA TTTATTGAAG GATCTGAAGC 3300

GATTGATAAG GAACATTTTA AGAAGAATCA TCTTCCTAAT TTAGCCAAAC AACTTGAAAA 3360

GTTTGGTTTT CCAACTTTTA ACTGTCAAGT CGAGAAGAAT GATGTCCTGA CCCAAGAGCA 3420

GGAAGAGGCC TTTCATGCTG AAAATGAGCA GATTGTTCAA GCTGCCAATG AGGAAGCGCT 3480

CCGTGCTATG GAACAACTGG AGCAGATGGC ACCTCCTCCA GCGGAAGAGA AACCAGCCTT 3540

TGATTTTCAA GCGAAAAAAG CTGCAGCTAA ACCCAAGCTG GATAAGGCGG AGATTACTCC 3600

TATGATCGAA GTGACGACAG AGGAAAATCG TCTGGTATTT GAAGGGGTTG TTTTTGATGT 3660

GGAGCAAAAA GTGACTAGAA CAGGTCGTGT TTTAATCAAC TTTAAAATGA CGGACTATAC 3720

TTCAAGTTTT TCTATGCAAA AGTGGGTTAA AAACGAGGAA GAGGCCCAGA AGTTTGACCT 3780

CATCAAGAAG AATTCTTGGC TCCGAGTTCG AGGGAATGTG GAGATGAATA ACTTCACACG 3840

CGATTTGACT ATGAACGTAC AGGATCTGCA GGAAGTTGTT CACTATGAGC GGAAGGATTT 3900

GATGCCAGAA GGTGAGCGTC GGGTTGAGTT TCATGCTCAT ACTAACATGT CGACTATGGA 3960

TGCTTTGCCA GAGGTCGAAG AGATTGTTGC AACAGCTGCT AAGTGGGGAC ACAAGGCGGT 4020

TGCTATCACG GACCATGGGA ATGTCCAGTC CTTTCCACAT GGCTATAAGG CGGCTAAGAA 4080

AGCGGGAATC CAGCTGATCT ATGGGATGGA AGCCAATATC GTGGAGGACC GTGTCCCTAT 4140

CGTCTATAAC GAAGTGGAGA TGGACTTGTC AGAAGCAACC TACGTGGTCT TTGACGTGGA 4200

AACGACGGGA CTTTCAGCTA TCTATAATGA CTTGATTCAG GTTGCGGCTT CTAAGATGTA 4260

CAAGGGGAAT GTTATTGCTG AATTTGATGA ATTTATCAAT CCTGGGCATC CCTTGTCAGC 4320

CTTTACTACA GAGTTAACTG GAATTACAGA TGATCATGTC AAAAATGCCA AACCACTAGA 4380

ACAAGTTTTG CAAGAATTCC AAGAATTTTG CAAGGATACG GTCCTAGTTG CCCACAATGC 4440

TACCTTTGAC GTTGGCTTTA TGAATGCTAA TTATGAGCGG CATGATCTTC CAAAGATTAG 4500

TCAGCCAGTT ATTGATACGC TGGAGTTTGC TAGAAACCTC TATCCTGAGT ATAAACGCCA 4560

TGGTTTGGGG CCTTTGACCA AGCGTTTTGG TGTGGCCTTG GAACATCACC ACATGGCCAA 4620

CTACGATGCG GAAGCGACTG GTCGTCTGCT TTTCATCTTT ATCAAAGAGG TAGCAGAAAA 4680

ACATGGTGTG ACCGATTTAG CTAGACTCAA CATTGATCTA ATCAGTCCAG ATTCTTACAA 4740

AAAAGCTCGG ATCAAGCATG CGACCATCTA TGTCAAGAAT CAGGTAGGTC TAAAAAATAT 4800

CTTTAAGCTG GTTTCCTTGT CTAATACCAA GTATTTTGAA GGAGTGCCAC GGATTCCGAG 4860

AACGGTTCTA GATGCCCATC GAGAGGGCTT GATTTTAGGT TCAGCCTGTT CAGAGGGTGA 4920

640

AGTTTTTGAC GTGGTCGTTT CTCAAGGTGT GGATGCGGCG GTTGAGGTGG CCAAGTATTA 4980

TGATTTTATC GAGGTCATGC CACCGGCTAT CTATGCACCC TTGATTGCCA AAGAGCAGGT 5040

CAAGGATATG GAGGAACTCC AGACCATTAT CAAGAGTTTG ATAGAGGTTG GAGACCGCCT 5100

TGGCAAGCCT GTTCTGGCTA CGGGAAATGT TCACTATATC GAACCGGAAG AAGAGATTTA 5160

TCGTGAAATT ATCGTCCGTA GTTTGGGACA GGGTGCGATG ATTAATCGAA CTATCGGTCA 5220

TGGTGAACAT GCCCAACCAG CACCACTTCC AAAGGCTCAT TTTCGAACGA CTAATGAGAT 5280

GTTGGATGAA TTTGCCTTTT TGGGAGAGGA ACTGGCTCGT AAACTGGTTA TTGAAAACAC 5340

CAATGCCTTG GCAGAAATAT TTGAATCCGT TGAAGTCGTT AAGGGTGACT TGTATACGCC 5400

TTTCATCGAC AAGGCTGAAG AAACAGTTGC TGAGTTGACC TATAAGAAAG CTTTTGAGAT 5460

TTATGGAAAT CCGCTGCCAG ATATTGTTGA TTTGCGGATT GAAAAAGAAT TAACATCCAT 5520

ACTGGGGAAT GGATTTGCTG TGATTTATCT GGCATCGCAG ATGCTGGTGC AACGTTCTAA 5580

TGAACGGGGT TATTTGGTTG GTTCTCGTGG GTCTGTCGGA TCTAGTTTCG TTGCGACCAT 5640

GATTGGGATT ACGGAGGTCA ATCCTCTCTC TCCTCACTAT GTCTGTGGTC AGTGTCAGTA 5700

CAGTGAGTTT ATCACAGATG GTTCGTACGG TTCAGGATTT GATATGCCCC ATAAGGACTG 5760

TCCAAACTGT GGTCACAAAC TCAGTAAAAA CGGACAGGAT ATTCCGTTTG AGACCTTCCT 5820

TGGTTTTGAT GGGGATAAGG TTCCTGATAT TGACTTGAAC TTCTCGGGAG AAGATCAGCC 5880

TAGCGCCCAC TTGGATGTGC GTGATATCTT TGGTGAAGAA TATGCCTTCC GTGCGGGAAC 5940

GGTTGGTACG GTAGCTGCCA AGACTGCCTA TGGATTTGTC AAAGGTTACG AGCGAGATTA 6000

TGGCAAGTTT TATCGTGATG CAGAAGTAGA ACGCCTCGCT CAAGGAGCGG CGGGTGTCAA 6060

GCGGACAACA GGCCAACACC CGGGGGGAAT CGTTGTTATT CCGAACTACA TGGATGTCTA 6120

CGATTTTACG CCTGTCCAGT ATCCAGCAGA TGATGTCACG GCTGAATGGC AGACCACTCA 6180

CTTTAACTTC CACGATATCG ATGAGAACGT CCTCAAACTC GATGTACTGG GACATGATGA 6240

TCCGACTATG ATTCGAAAAC TTCAGGATTT GTCTGGTATT GACCCTAATA AAATTCCTAT 6300

GGATGACGAA GGCGTGATGG CACTCTTTTC TGGGACTGAT GTGCTAGGGG TAACACCTGA 6360

ACAAATTGGA ACGCCTACGG GTATGTTGGG GATTCCAGAG TTTGGAACAA ATTTCGTACG 6420

TGGAATGGTA GACGAAACCC ATCCGACAAC CTTTGCGGAA TTGCTTCAGC TGTCTGGTCT 6480

GTCCCACGGT ACTGATGTTT GGTTGGGGAA TGCTCAGGAT CTGATTAAGC AAGGAATAGC 6540

GGACCTATCG ACTGTTATCG GTTGTCGGGA CGACATCATG GTTTACCTCA TGCATGCGGG 6600

TCTGGAACCT AAGATGGCCT TTACCATTAT GGAACGGGTA CGTAAGGGTT TGTGGCTAAA 6660

GATTTCAGAA GAGGAGAGAA ATGGCTATAT CGAAGCAATG AAGGCTAATA AGGTGCCAGA 6720

GTGGTATATC GAATCCTGTG GGAAAATTAA GTACATGTTC CCTAAGGCCC ATGCGGCAGC 6780

CTACGTTATG ATGGCCTTGC GTGTAGCTTA CTTCAAGGTT CACCATCCTA TTTATTACTA 6840

CTGTGCTTAC TTCTCCATTC GTGCTAAGGC TTTTGATATC AAGACCATGG GTGCGGGCTT 6900

GGAGGTCATC AAGCGCAGAA TGGAAGAAAT CTCTGAAAAA CGGAAGAACA ATGAAGCCTC 6960

TAATGTGGAA ATCGATCTCT ATACAACTCT TGAGATTGTC AATGAGATGT GGGAACGAGG 7020

TTTCAAGTTT GGTAAATTAG ATCTCTACTG TAGTCAGGCG ACAGAGTTCC TCATCGACGG 7080

GGATACCCTT ATCCCACCAT TTGTAGCAAT GGATGGTCTG GGAGAGAACG TTGCCAAGCA 7140

ACTGGTGCGG GCGCGTGAAG AGGGAGAATT CCTCTCTAAA ACAGAACTAC GCAAGCGTGG 7200

TGGACTCTCA TCAACCTTGG TTGAAAAGAT GGATGAGATG GGTATTCTTG GAAATATGCC 7260

AGAGGATAAC CAGTTGAGTT TGTTTGATGA GTTGTTTTAA AAAATTGCTT AATAATCTAT 7320

TAAAAGAGGC TAACGTATAT CCAATAGATT TACATTAGCT TTCTTTTTTG TTAAAATAGT 7380

CTATGGAAAG AGGGTGAGAG TATGTCAAAG ATGAGTATAA GCATCCGTCT GGATAGTGAG 7440

GTTAAGGAGC AGGCCCAACA GGTGTTTAGT AATCTGGGAA TGGATATGAC AACAGCTATT 7500

AATATTTTCC TTCGTCAGGC AATTCAATAT CAGGGATTAC CTTTTGATGT TAGACTAGAC 7560

GAAAATCGGA AGTTGCTCCA AGCGTTAACG GATTTAGACC AAAATCGTAA TATGAGCCAG 7620

TCTTTTGAAT CAGTCTCAGA TTTGATGGAG GACTTACGTG CTTAAGATTC GTTATCATAA 7680

ACAGTTTAAA AAAGATTTTA AGTTGGCTAT GAAGCGTGGT TTGAAGGCAG AATTATTAGA 7740

AGAAGTTTTG AATTTTCTGG TTCAAGAAAA AGAACATCCT GCCAGAAATC GTGATCATTC 7800

ATTGACGGCA TCCAAGCATT TTCAAGGAGT TCGTGAATGC CATACCCAGC CAGATTGGCT 7860

TTTGGTTTAT AAAGTAGACA AGTCGGAATT GATTTTAAAT TTGCTGAGGA CAGGCAGTCA 7920

CAGTGATTTA TTTTAATCTA TTTTAAGGGG GTTCTCATGA AACTAAGAAT ATTTGCGGAA 7980

GATAAGCCGG CTAAGAAGGT ATTTGAATAT CAATTAGAAC TTGCTGATCG TACAATTCTT 8040

CTATCGACAG CACTCTTGTC AGGTGCTATT GCTTTAGCAG GAATCTTTTC TGCTTTGAAA 8100

GAAAAATAAA AATAGAAAAG AGAAAACAGA ATGGTTTTAC CAAATTTTAA AGAAAATCTA 8160

GAAAAATATG CGAAATTGTT GGTTGCGAAC GGAATTAACG TGCAACCTGG TCACACTTTG 8220

GCTCTCTCTA TTGATGTGGA GCAACGTGAA TTGGCACATC TAATCGTGAA AGAAGCTTAT 8280

GCCTTGGGTG CGCATGAGGT CATCGTTCAG TGGACAGATG ATGTGATTAA CCGTGAGAAA 8340

TTCCTCCATG CCCCGATGGA GCGTTTGGAC AATGTGCCAG AATACAAGAT TGCTGAGATG 8400

AACTATCTCT TGGAGAATAA GGCTAGCCGT CTTGGAGTTC GTTCATCTGA TCCAGGTGCC 8460

TTGAACGGAG TGGACGCTGA CAAGCTTTCA GCTTCTGCTA AAGCTATGGG ACTTGCCATG 8520

AAGCCTATGC GTATCGCAAC TCAATCTAAC AAGGTTAGCT GGACTGTAGC AGCTGCAGCA 8580

GGACTTGAGT GGGCTAAGAA AGTCTTCCCA AATGCTGCGA GCGACGAAGA AGCAGTTGAT 8640

TTCCTTTGGG ACCAAATTTT CAAAACTTGC CGTGTCTACG AAGCAGATCC TGTTAAGGCT 8700

TGGGAGGAAC ATGCAGCCAT TCTCAAGAGC AAGGCCGATA TGCTTAATAA GGAGCAATTT 8760

TCAGCCCTTC ACTACACAGC GCCAGGAACA GATTTAACAC TTGGTTTGCC AAAGAACCAC 8820

GTTTGGGAAT CAGCTGGTGC TGTCAATGCA CAGGGCGAAG AATTCTTGCC AAATATGCCA 8880

ACAGAAGAGG TCTTCACAGC GCCTGACTTC CGTCGTGCAG ATGGTTATGT CACTTCTACA 8940

AAACCGCTTA GCTACAACGG AAATATCATT GAAGGCATTA AGGTGACCTT TAAGGATGGA 9000

CAAATCGTAG ATATCACTGC TGAGAAGGGT GATCAGGTTA TGAAAGACCT TGTCTTTGAA 9060

AATGCGGGTG CGCGTGCCTT GGGTGAATGT GCCTTGGTAC CAGATCCAAG TCCAATTTCT 9120

CAGTCAGGCA TTACCTTCTT TAACACCCTT TTCGATGAAA ATGCGTCAAA CCACTTGGCT 9180

ATCGGTGCAG CCTATGCGAC TAGCGTTGTT GATGGAGCGG AGATGAGCGA AGAGGAGCTT 9240

GAAGCTGCAG GGCTTAACCG TTCAGATGTT CACGTAGACT TTATGATTGG TTCTAACCAA 9300

ATGGATATCG ATGGTATTCG TGAGGATGGA ACGCGGGTAC CTCTTTTCCG TAATGGGAAT 9360

TGGGCAAATT AAGGAGATAA TATGTTAGGA AGTATGTTCG TTGGTCTCCT AGTGGGATTT 9420

TTAGCAGGTG CTATGACCAA TCGTGGAGAG CGAATGGGAT GTTTTGGAAA AATGTTTCTC 9480

GGTTGGATCG GAGCCTTTCT AGGTCACTTG CTCTTTGGAA CTTGGGGGCC AGTTTTATCA 9540

GGAACAGCTA TTATCCCAGC GATTTTAGGA GCCATGATTG TTTTAGCTAT TTTTTGGAGA 9600

CGAGGAA 9607

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14231 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

CTACAAGATA ATTCCAGCTA TAACATCCGC TATAATAGTA AGAGCGAGCT CTATGATAAG 60

GCTCATTAGT TTCACCTCCT CTCACGAACC CATAGGAACG TAATCGGTAA CCGATGACAA 120

AAATAGTATA CCACAATACA TTTAGATCAT CAAGGTCACT TAATTCTTGA AATATCAGAT 180

CTAAGAGAAA AATCTTTAAA ATCAGAAAAA CGCATAATAT CAGGTGTGCA AAAACTTGAT 240

ACTATGCGTT TTATTGTGGG AAGGTTTACT CCATTTTCTC CTGAAATTGA GTTTTTGTCC 300

AGCCTCTGTT TTTAGGGTTG CTAAGAAAAT AATGTCATGT GGTGAATATT TGTAAATCAG 360

TCAGCAGACA GAACGATACT CTTCGAAAAT CTCTTCACAT CATGTCAGCT TCGTCTTTCC 420

GTATATATGT GACTGACTTC ATCAGTTCTA TCTACAACCT CAAAACAGTG TTTCGAGCTG 480

ACTTGATCAA TTTTCAAATC TGTACTTTGA GCAAGCTGAG ACTAGCTTCC TATTTGATTT 540

TCATTGAATA TCAGAAACCC ATTCTCCATC AAATAATTCG ACTGCGTCTA ATAATTTTTG 600

ATCTGGCACG GTGTCTGAAA TAAAGGTTGT GTATTTGGAG AGGGGATTAA TTTTAAAAAA 660

TCCAGTCTTG TAAAATTTAG AACTATCAAT CAGTAAGATG GTTTCATGGG CTTTGTCAAT 720

AATATTCTTT TTTGAAATAG CTTGGCTGAG AGAAGCTTCA TAAACATATT GGTCATCAAT 780

ACCTCTTGCT GAACAAAATG CTAAATCGAT ATTAAAATGA TCTAATAAAG AATTTTCCTT 840

ATCATAGTTG ACCACGGAAC AGGATTGATG TTTGACCTCG CCAGATGTGA TAAAGATTTT 900

GGAGCTATCT TTAACAGTTT CAGATAGGGT TTGTGCAGTA TGTAAACCAT TTGTAAAAAT 960

AATCAAATTA TCAAGTTCAG AAAGATAGGG ACAGAGTTCG TAGACAGTAG TACTAGAATC 1020

TAGATAGATA CACATACCAG ACCGAATAAA GTCTTTAGCG AGACTAGCGA TTAGTCTTTT 1080

TTGCCTAGTA CTTTCTCCTT CACGTATTTG ATGAGAAAGT TCAATTGTGT TCATAGAGGA 1140

CAGGGTCACG TATCCGTGCT TTCTTTTGAT AAGACCTTGA TTTTCTAAGA AAATTAAATC 1200

ACGACGTAAG GTACTTGTGC TGGAGAAAGT GATTTCTGCC AGCTCTTTTA CGGCAATTCT 1260

TTTTTTCTTT TTGATAATTT CAATCAATTC AAGTACACGT TCATCTTTTA TCATAAGCTC 1320

CTCCTAATTT ATCATTTCAA CTATATTATA GCACAAATTG GAGGAATTTG AATTATTTTT 1380

ATGAATATTG GGTTAACATT TGAACATTAT TCAAGTAAGC GTTCACATAT TGAAAAAATA 1440

AAACGTGGGG ATTATAATAA AGTTAATCmA GGACGAAGAG AGAAGAAAAA TGGAAGCGGT 1500

TTTAGCAATA GATTTAGGTG CGACTTCTGG AAGAGCAATC GTTGGTTACC TTTCTGAAAA 1560

TAAACTAGTA ATGGAAGAAA TAAATCGCTT TTCTAATCTA CCTATTAGAG TAAAAGGGCA 1620

TTTATCTTGG GATATTGACT TTCTACTAGC TAAAATTCTT GAAAGTATCC GCTTGGCTAA 1680

TACTAGTTAC AAGATTTTAT CTATCGGTAT TGACACATGG GGAGTTGATT TTGGACTGAT 1740

TGATAATGAA GGTAAGCTGT TATTACAACC TGTTCATTAT CGTGATGAAA GAACAAAGGG 1800

AGTGTTAAAG GAAATATCTG AAATGACTGA ATTAGAAAAA CTGTATTCAG AGACAGGAAA 1860

TCAGATTATG GAGATAAATA CCTTGTTTCA ACTCTTTAAG GCACGTCAAG AATCTCCTGA 1920

CTCTTTCTAT AAGACCAATA AGATTCTTTT AATGCCAGAT TTGTTTAATT ATCTCTTGAC 1980

AGGTAAGTTT GCTACAGAAA AAAGCATTGC TTCAACAACT CAATTATTTG ATCCTAGGAG 2040

TCAAAATTGG AATCAGAATA TCTTAAAACT ATTTGAATTG GATTCATCTT TACTTCCTGA 2100

AATTGTTTCA GAGGGAAATG TTCTTGGAAG GATAAAAGAG GAGTATGGTT TAGGCGATAT 2160

TCCTGTTGTG AATGTTTGTA GTCATGATAC AGCAAGCGCG ATTGTCTCAG TACCTAAGAC 2220

AGAAGGTAGT TTATTTATTT CATCAGGTAC TTGGTCTTTG GTTGGAGTGG AACTTACTTC 2280

ACCGATTCTT ACTACCGAAT CCTTCAGTTA TGGATTTACA AATGAAGTCG GTAAAGATGG 2340

AGTGATTACA TTTCTGAAGA ATTGTACAGG GTTGTGGATC ATAGAGGAAC TAAGACGTTC 2400

ATTTGAACGA AGAGGGAAAG CCTATTCTTT TGATGATATT AGGACAATGG TGGAGAAAGA 2460

AAAAGAAAAT CTTCCTCTGA TTGATACTGA ATCAACTGAA TTTGCAACAG AATCTGATAT 2520

GCACAAGACT TTGACAGAAT ATCTAGCTTA TCATCATGAA ACTAGAGAGT GGACAGATGG 2580

ACAACTATTT AAGATTGTTT ATGAAAGCCT AGCTGAAACG TATAGGAAAG CGATAGAGTT 2640

ACTAGAAGAA CTAACTCATA AGGTTTATAA GAGGATATAT GTGATTGGAG GAGGTGCTAG 2700

AGCCAGTTAC TTTAACCAAA TGATTGCTGA TAGAACTGGT AAAGAGGTTC TTACAGGTTT 2760

GACTGAGGCT ACAGCTGTGG GGAATATTGT TGTGCAGCTC ATAGCTATGG GACAATTAAA 2820

AGGGATGGAA GAGGCTCACC ATGTTATTGA GGAGTTTCTA CAATTAGAGA GTTATTACTC 2880

CCAAAAGAAT TAAAAAGATT GAGAGTTTGT AAATTTGCCT CCCTCCCCCT TCTTAGCTTT 2940

TGTGCAGGAA GGGGGGATAA TTGGTGAATT GAAAAATATT TAGTGTTTTG ATATGAGGAG 3000

GACAAGGATG TCAGATGTAA AACAAGAATT AATTAAATAT GGTAAGAAGC TAGTAGAAAC 3060

AGATTTGACG AAAGGAACAG GTGGGAATCT CAGCGTTTTC GATCGTGAAA AACAATTGAT 3120

GGCAATTACC CCGTCGGGTA TTGATTTCTT TGAAATCAAA GAATCCGATA TTGTAGTGAT 3180

GGATATTAAT GGAAATGTTG TAGAGGGAGA ACGCTTGCCA TCTAGCGAAT GGTATATGCA 3240

TTTGATTCAA TATCAAACTC GTGATGATAT CGATGCAATT ATCCATGCTC ATACAACTTA 3300

TGCAACAGTA TTAGCTTGTC TCAGAGAACC ACTTCCAGCG AGTCATTATA TGATTGCAGT 3360

GGCAGGGAAA GATGTTCGGG TAGCTGAGTA TGCAACATAT GGCACGAAAG AATTGGCTGT 3420

GAATGCAGCT AAAGCAATGG AAGGTCGTAG AGCAGTTTTA CTAGCGAATC ATGGAATTTT 3480

AGCAGGTGCA CAAAATTTAT TGAATGCATT TAATATTGTT GAAGAAGTTG AATATTGTGC 3540

AAAAATTTAT TGTTTAGCTA AGAATTTTGG AGAGCCAGTA GTTCTTCCTG ATGAGGAGAT 3600

GGAATTGATG GCAGAAAAAT TTAAAACATA CGGTCAGAGA AAATAGGGAG GATATTAATG 3660

TTAAAACATA TACCGAAAAA TATTTCTCCA GATTTATTGA AGACTTTAAT GGAAATGGGA 3720

CATGGAGATG AAATAGTATT AGCTGACGCG AATTATCCTT CTGCCTCATG TGCAAATAAG 3780

CTAATTCGTT GTGATGGTGT AAATATTCCA GAATTATTAG ATTCCATTCT GTATTTAATG 3840

CCATTAGATA GTTACGTCGA TAGTTCAATT CAGTTTATGA ACGTTGTTTC GGGTGATGAT 3900

ATTCCTAAGA TATGGGGTAC CTATAGACAG ATGATTGAAG GTCATGGTAC AGATCTTAAA 3960

ACGATTACTT ATCTTAGAAG AGAAGACTTT TATGAACGTA GTAAGAAAGC TTATGCTATT 4020

GTTGCTACAG GAGAAACTTC ACTTTATGCT AATATTATCC TTAAGAAAGG AGTAGTTGTT 4080

GAAAGAGAAA ATGTTCAATA GAGGAATTTT AGTTGCCAGT CATGGTAATT TTGCTAGCGG 4140

AGCTCTCATG ACCGCAGAAA TGTTTGTTGG TGAGACAACA AATGATAGAG TTAGGACATT 4200

AGGTTTGATG CCTGGAGAGA ATATTGTAGA GTTTGAGCAT TATTTTAAAA ATCAAGTGGA 4260

TGAACTGTTA GACTCAAATC AAGAGGTTAT CGTTTTGACT GACTTGATTG GAGGAAGTCC 4320

TAATAATGTG GCTTTGTCAC GGTTTTTAAA TTTGGATTCA GTTGATATTG TAACAGGGTT 4380

TAATATCCCT CTCCTAGTGG AATTAATATC AAGTTATGAT TCAAAAATCA ATTTAGAAGA 4440

AATTGTTCAC AATGCTCAAA ATAGTTTGTT TAATGTTAAA CAACAACTTA ACGTAGAGGA 4500

GGAAGAAGAT TTATGTCTAT AGAGTTTGTT CGTATTGATG ACCGTCTGGT ACATGGTCAA 4560

GTTGTCACTA CGTGGCTAAA AAAGTATGAT ATTGAGCAAG TTATCATTGT TAATGATCGC 4620

ATCTCAGAAG ATAAAACACG ACAATCTATT TTAAAGATTT CTGCACCGGT AGGTTTAAAA 4680

ATTGTTTTCT TTAGTGTAAA ACGGTTTGTG GAAGTTTTAA ACTCTGTGCC AATAAAAAAG 4740

AGAACAATGC TGATATATAC AAATCCAAAA GATGTGTATG ATTCTATTGA AGGAAATTTA 4800

AAATTGGAGT ACCTCAATGT AGGACAGATG AGTAAAACGG AGGAAAATGA AAAGGTAACG 4860

GGAGGTGTAG CTCTAGGTGA AGAAGACAAA TATTATTTTA AGAAAATAGT TGATAAGGGA 4920

ACGAGAGTTG AAATTCAAAT GGTTCCTAAT GATAAAGTTA CAATGTTGGA AAAATTTTTA 4980

TAAAAATAAT TTAAGGAGGT ACAGTATATG CTATTCACAC AAGCATTACT GGTGACATTA 5040

GTTGGGATTA TTGCCACTAT TGACTATAAT GGACCGTTAT TTATGATTCA CCGTCCGTTA 5100

GTTACAAGTG CAATGGTTGG CTTAGTATTA GGAGATTTCA CCCAAGGTGT TCTTATTGGT 5160

TCAGCTCTTG AATTAACTTG GCTCGGTGTA ACAGGTATTG GAGGTTATAC TCCACCAGAT 5220

ACTATTTCAG GTGCGATTAT TGGTACTGCA TTTGGTATTT TATCTGGTCA AGGAGAAACT 5280

GCTGGTATCG CTATAGCAGT TCCAATTGCA GTTGCTACCC AACAGTTGGA TGTTCTTGCA 5340

AAAACTTTAG ATGTTTATTT TGTGAAAAAA GCTGATAATG ATGCTAAAAA CGGAGATTAT 5400

TCAAAGATCG GTTTTTATCA TTATTCAAGT TTGGTTTTAA TCACGTTATT TAAAATTGTA 5460

CCAATTTTCC TAGCTATTAT GCTTGGAGGG GAATATGTGG CAGACTTGTT TGCTAAGGTT 5520

CCACCAATCG TTATGCAGGG ACTTAACTCT GCAGGTGCTT TACTACCTTC AATTGGTTTT 5580

GGTATGCTTT TAAATATGAT GCTCAAGAAA AATATGTGGG TATTCTTGTT GATTGGATTC 5640

ATTTGTTCTG TGTATGGAGG AATGTCAACC ATTGGGATCT CACTAGTTGG TATTGCGGTA 5700

GCATACTTCT ACGATATGAT TGGAAGCAAA CCACAAGAAA CAACTTCAAG TAGTGATGTT 5760

GAGGAGGATC TTGATCTATG ATGAATAATA AAGTAACTAA AGTTGAACTT AAAAAAGTTT 5820

TCAAACGAAG TTTTATGTAT GGTTCTTCAT GGAACTATGA GAGAATGCAG AACCTAGGTT 5880

TTCTATATAC AATTCTTCCA GTATTGAAAA AACTATACCC AGACAAAGAT TCAGCTTCTC 5940

CTGCAATGAA ACGTCACCTT GAGTTTTTCA ATACTCATCA AACAGCGGCA CCATTTATTC 6000

TTGGAGTTAC TTCCGCTATG GAAGAACAAG AAGGAAATGA AGGTGCAGCT TCAATTACTG 6060

GTATTAAAGT TGGCTTGATG GGGCCACTGG CTGGTCTAGG AGATAGTTTG TTCTGGCTGA 6120

CACTAGTTCC TATCTGTTTT AGTATTGGTG CGTCTTATTC TAAAGACGGC GGTGCTTTAG 6180

GTATCTTTAT CGCCTTAATA TTGTTTAATA TTATTAATAT TCCTGTTAAA TATTTCGGTT 6240

TGAAATATGG GTATACTAAG GGTTCTAGTC TTATCCAAGA AAATAATACA AAAGGAACAT 6300

TGAATCGCGT TACGAGTATG GCGACAGCAT TAGGGCTAGT ACTAGTGGGT GGTTTGATTC 6360

CATCAATGGT TGGTATTAAT TTTGGATTAG AATTTAAGCA GGGGGAACTT GTTATTTCTG 6420

TTCAAGAAAT GATTACAAAA TTAATTCCAG GATTTATCCC TATGGCTTTG ACTTTATTAA 6480

TGTGTAAATT AATTAGAAAA GGAAAGAATC CGGTTGTACT AATCTTTAGT GTTATGGCTA 6540

TTGGAGTTAT TCTAGTTGTT TTAGGAATTT TGAAGTAGTA GAAAGTGTGG AGGTGGTATT 6600

TGGGATATCA CCTCCATTTT GGAAGAGAGG TAAAGAGTGA AATTATGGTA TAAGAAAGCT 6660

GCCGCAAATT GGAATGAAGC CTTGCCGATT GGGAACGGTC ATTTAGGTGG TATGATTTAT 6720

GGTTCAGCTA CAAAAGAATG TATTCAACTA AACGATGAGA CTATTTGGTA TAGAGGAAAG 6780

TCAGATAGAA ATAATCCAGA CTCACTATTG CATCTTAAAA AAATTCGGGA ATATCTTTTA 6840

GATGGAGAAA TTCAGAAAGC CGAAGAATTG ATAAAGTTAA CAGTGTTTGC TACCCCAAGA 6900

GATCAAAGCC ACTATGAATT ACTTGGGGAA CTTTACATTG AGCATATAGA TATTCAGTCT 6960

TGTGCTCTTT CATTGTATGA AAGAGAGCTA GATTTAGATA CAGCTATTTC TAATGTTGTG 7020

TTTGAGCCTA ATAGTTGTAA TTTACAAATA AAAAGAGAAT ATTTTACGAG TTTTAATAAG 7080

AATATTTTAT GTTGCCGTAT AGTGTCATCA GTTCAAAACA CATTAAATTT AAACATTAAT 7140

TTGGGTAGAA ATAAACGGTT TAATGACGAA GTATCTAAAC TGGATTCAAG TACAATTTTA 7200

ATGTCGGCCT CTGCTGGAGG TAGAAAAGGT GTTCAGTTTA AAGTAGTATG TCATTCTAAG 7260

GTTACGGATG GTGAAGTAAG TGTATTGGGA GAGACAATAG TTATTCGGAA TGCTACAGAG 7320

GTATTTCTTT ATCTCAAATC AATGACGGAT TATTGGGGAA ATATAGATAT TTCTTCTCTT 7380

CAGGGAGAAT TTAGTAGTAT TGATTACTTT ACAGAAAAAG ATGAACATGT AAAAAAATAT 7440

CAGGAGCAAT TTAATAGAGT TGATTTTAAA CTAGACTATA GTAAAGGTTG TCTTAGCATT 7500

CCAACGAATC TACTTCTTGA AAACACTAAA AAGTATAGTA ACTACTTGAC TAACTTGTTA 7560

TTTCATTATG GAAGATATCT GTTAATATCG TCTAGTCAAC CGAATGGTTT ACCTGCCAAT 7620

CTTCAAGGAA TATGGTGTGA TGAATTAAAT CCAATTTGGG GTTCTAAATA TACGATTAAT 7680

ATTAATACTC AAATGAATTA TTGGATGGTA GGTCCATGTG ATTTACCAGA AGTAGAATAT 7740

CCATTATTTG ATATGCTCGA AAGAATGAGA GAACCGGGAA GACTAACCGC TAAGAAAATG 7800

TATGGAGCTA GAGGTTTTAC AGCACATCAT AATACGGATG GTTTTGGCGA TACGGCTCCC 7860

CAATCTCATG CCATGGGGGC TGCAATTTGG GTATTAACTA TTCCATGGTT ATGTACTCAT 7920

ATTTGGGAAC ACTATTTATA TTTCCAAGAT GAGCGTATTC TTACGGAACA TTTTGAAATG 7980

ATAAAAGAAG CATTTCTTTT CTTTGAAGAT TATTTATTTG AGGTGGATGG CTACTTGATG 8040

ACAGGTCCAA GTGTCTCACC GGAAAATAAA TATCGCTTAA AAAATGGTAT TGAAGGAAAT 8100

GCTTGTCTAT CATCTACAAT TGATAATCAA ATTCTAAGAT ATTTTTGTGA TTCATGCATT 8160

GGCATTGCAA AACAATTAGG AGACAATTCG GATTTTATTA GTCGTGTGAA GGAGTTAAAA 8220

AAGAAACTAC CTAAAACAAA AATAGGTAGT AATGGGCAAA TCCAAGAATG GTTAGAAGAT 8280

TATGAAGAAG TAGAGCCTGG GCATAGACAC ATTTCACCTC TATTTGGGCT TTATCCTTAT 8340

AATGAGATTG ATATTCATAA AACTCCGGAA TTAGCAGAAG CAGCTAAAAT CACTATCAAT 8400

AGGAGATTAT CAAACGCTAA TTTTTTATCT TCACAGGAGA GGGAGCAAGC GATTAATAAT 8460

TGGTTAGTAA GTGGTTTGCA TGCTAGTACA CAAACAGGTT GGAGTGCTGC ATGGCTGATT 8520

CATTTTTTTG CGAGACTATA TCAAGGTGAA CCTGCTTATA ACCAGATTAA TGGTTTGTTA 8580

AATAATGCGA CTCTTGGCAA TTTATTTCTT GACCATCCAC CATTTCAAAT TGATGGTAAT 8640

TTAGGTTTGG TGAGTGGAAT TTGTGAATTA TTAGTACAGA GCCATCATAA TTGGTTATCA 8700

CTAATTCCAG CTTTACCTTC TGCTTGGTCA GAAGGAGAAG TGAAAGGTTT CAGAGTAAGA 8760

GGAGGATATA AGGTATCGTT TGCTTGGAAA AATGGGGATA TAACATTCCT AAAATTGGAA 8820

GGAGGAAACA AAGATCAAAA AGTAAGAGTA AGAATATATG GCAAAAATAC TGATGTACAA 8880

AATATTGAAT TGGTATTTAA TTCAGAAAAA ATTATTGAGT TAAATTTTTA GGTATAAGTC 8940

ATGAATAAAG AAAAAATAAA AAGAAAATTA ATCACAATAT TGTTTGTATG TATTGGGATG 9000

TTATGTTTTG GATTGTTAGC AGGAGTTAAG GCTGATAATC GTGTTCAAAT GAGAACGACG 9060

ATTAATAATG AATCGCCATT GTTGCTTTCT CCGTTGTATG GCAATGATAA TGGTAACGGA 9120

TTATGGTGGG GGAACACATT GAAGGGAGCA TGGGAAGCTA TTCCTGAAGA TGTAAAGCCA 9180

TATGCAGCGA TTGAACTTCA TCCTGCAAAA GTCTGTAAAC CAACAAGTTG TATTCCACGA 9240

GATACGAAAG AATTGAGAGA ATGGTATGTC AAGATGTTGG AGGAAGCTCA AAGTCTAAAC 9300

ATTCCAGTTT TCTTGGTTAT TATGTCGGCT GGAGAGCGTA ATACAGTTCC TCCAGAGTGG 9360

TTAGATGAAC AATTCCAAAA GTATAGTGTG TTAAAAGGTG TTTTAAATAT TGAGAATTAT 9420

TGGATTTACA ATAACCAGTT AGCTCCGCAT AGTGCTAAAT ATTTGGAAGT TTGTGCCAAA 9480

TATGGAGCGC ATTTTATCTG GCATGATCAT GAAAAATGGT TCTGGGAAAC TATTATGAAT 9540

GATCCGACAT TCTTTGAAGC GAGTCAAAAA TATCATAAAA ATTTGGTGTT GGCAACTAAA 9600

AATACGCCAA TAAGAGATGA TGCGGGTACA GATTCTATCG TTAGTGGATT TTGGTTGAGT 9660

GGCTTATGTG ATAACTGGGG CTCATCAACA GATACATGGA AATGGTGGGA AAAACATTAT 9720

ACAAACACAT TTGAAACTGG AAGAGCTAGG GATATGAGAT CCTATGCATC GGAACCAGAA 9780

TCAATGATTG CTATGGAAAT GATGAATGTA TATACTGGGG GAGGCACAGT TTATAATTTC 9840

GAATGTGCCG CGTATACATT TATGACAAAT GATGTACCAA CTCCAGCATT TACTAAAGGT 9900

ATTATTCCTT TCTTTAGACA TGCTATACAA AATCCAGCTC CAAGTAAGGA AGAAGTTGTA 9960

AATAGAACAA AAGCTGTATT TTGGAATGGA GAAGGTAGGA TTAGTTCATT AAACGGATTT 10020

TATCAAGGAC TTTATTCGAA TGATGAAACA ATGCCTTTAT ATAATAATGG GAGATATCAT 10080

ATTCTTCCTG TAATACATGA GAAAATTGAT AAGGAAAAGA TTTCATCTAT ATTCCCTAAT 10140

GCAAAAATTT TGACTAAAAA TAGTGAGGAA TTGTCTAGTA AAGTCAACTA TTTAAACTCG 10200

CTTTATCCAA AACTTTATGA AGGAGATGGG TATGCTCAGC GTGTAGGTAA TTCCTGGTAT 10260

ATTTATAATA GTAATGCTAA TATCAATAAA AATCAGCAAG TAATGTTGCC TATGTATACT 10320

AATAATACAA AGTCGTTATC GTTAGATTTG ACGCCACATA CTTACGCTGT TGTTAAAGAA 10380

AATCCAAATA ATTTACATAT TTTATTGAAT AATTACAGGA CAGATAAGAC AGCTATGTGG 10440

GCATTATCAG GAAATTTTGA TGCATCAAAA AGTTGGAAGA AAGAAGAATT AGAGTTAGCG 10500

AACTGGATAA GCAAAAATTA TTCCATCAAT CCTGTAGATA ATGACTTTAG GACAACAACA 10560

CTTACATTAA AAGGGCATAC TGGTCATAAA CCTCAGATAA ATATAAGTGG CGATAAAAAT 10620

CATTATACTT ATACAGAAAA TTGGGATGAG AATACCCATG TTTATACCAT TACGGTTAAT 10680

CATAATGGAA TGGTAGAGAT GTCTATAAAT ACTGAGGGGA CAGGTCCAGT CTCTTTCCCA 10740

ACACCAGATA AATTTAATGA TGGTAATTTG AATATAGCAT ATGCAAAACC AACAACACAA 10800

AGTTCTGTAG ATTACAATGG AGACCCTAAT AGAGCTGTGG ATGGTAACAG AAATGGTAAT 10860

TTTAACTCTG GTTCGGTAAC ACACACTAGG GCAGATAATC CCTCTTGGTG GGAAGTCGAT 10920

TTGAAAAAAA TGGATAAAGT TGGGCTTGTT AAAATTTATA ATCGCACAGA TGCTGAGACT 10980

CAACGTCTAT CTAATTTTGA TGTGATTCTA TATGACAATA ATAGAAACGA AGTTGCTAAG 11040

AAACATGTTA ATAATTTGTC GGGTGAATCT GTTAGTCTAG ATTTCAAAGA AAAAGGAGCA 11100

AGGTATATTA AAGTTAAATT ACTAACGAGT GGAGTGCCTT TGAGTTTAGC AGAAGTAGAG 11160

GTTTTTAGAG AATCAGATGG TAAGCAATCT GAAGAGGATA TAGATAAAAT AACAGAAGAT 11220

AAAGTAGTCT CTACAAATAA GGTAGCTACT CAAAGTTCAA CCAATTATGA GGGTGTAGCT 11280

GCTTTAGCAG TTGATGGTAA TAAAGATGGA GATTACGGAC ATCATTCGGT GACTCATACT 11340

AAGGCAGATT CTAACGCTTG GTGGCAGGTC GATCTGGGAG AAGAGTTTAC GGTTTCTAAA 11400

GTTGATATTT ATAATAGAAC AGATGCCGAA CCTCAGCGTT TATCTAATTT TGATGTTATT 11460

TTTCTATCTT CATCAGGAGA AGAAGTTTTT AGAAGACATT TTGATAAAGT AGTTGATGGT 11520

TTGTTATCTT TAAAAGTACC TTCTGTAGGG GCTAAGCTAG TCAAAATAGA ATTAAAATCA 11580

GCAGCTATTC CGTTAAGTTT AGCGGAAGTT GAAGTCTATG GTTCAAAGAG AACTCCGAAG 11640

AAACTTTCTA ATATTGCATT AACAAAAGAA ACTCGACAGA GTTCAACGGA TTACAATGGT 11700

TTTTCTCGTC TAGCAGTTGA TGGAAATAAA AACGGAGATT ATGGTCATCA TTCAGTGACT 11760

CATACCAAAG AAGATTCTCC TTCATGGTGG GAGATAGATT TAGCACAAAC CGAAGAATTA 11820

GAAAAGTTAA TTATTTATAA TAGAACAGAT GCTGAAATTC AGAGATTATC AAATTTTGAT 11880

ATTATTATAT ATGATTCAAA TGATTATGAA GTTTTTACAC AACATATTGA CAGTTTAGAA 11940

AGCAATAATC TATCCATAGA CTTAAAAGGA CTGAAGGGAA AAAAGGTTAG AATTTCTTTG 12000

AGAAGCGCAG GAATTCCTTT AAGTTTAGCA GAGGTAGAGG TTTATACTTA TAAGTAATTT 12060

TAAAAATTAT CACCCAGGCT ACCGTAAATA TAATGGAGAT GGTAGTATGA AAGAAACAGA 12120

AAAATAAGAG GAAAATAGTA TGATTCAACA TCCACGTATT GGGATTCGTC CGACTATTGA 12180

TGGTCGTCGT CAAGGTGTAC GCGAATCACT TGAAGTGCAA ACAATGAACA TGGCTAAAAG 12240

TGTGGCAGAT TTGATTTCAA GCACATTGAA ATATCCAGAT GGGGAACCTG TGGAATGCGT 12300

GATTTCTCCA TCTACTATTG GCCGTGTACC AGAGGCTGCA GCTTCCCATG AGTTGTTTAA 12360

AAAATCAAAT GTTTGCGCAA CAATTACAGT TACACCATGC TGGTGTTATG GTAGTGAAAC 12420

TATGGATATG TCTCCAGATA TTCCTCATGC TATTTGGGGA TTTAATGGGA CAGAACGCCC 12480

AGGAGCTGTC TATCTTGCAG CTGTACTAGC TTCACATGCT CAAAAAGGGA TTCCAGCCTT 12540

TGGGATTTAT GGAAGAGATG TTCAGGAAGC TAGTGACACA GATATTCCAG AAGATGTCAA 12600

AGAAAAACTT TTACGCTATG CGCGTGCAGC TCTTGCAACT GGCTTGATGA GAGACACTGC 12660

TTACCTATCA ATGGGTAGTG TTTCGATGGG GATTGGTGGT TCTATTGTAA ATCCGGATTT 12720

CTTCCAAGAA TACTTAGGAA TGCGAAATGA ATCGGTAGAT ATGACGGAGT TCACGCGCCG 12780

TATGGACCGT GGTATTTACG ACCCTGAAGA GTTCGAACGT GCGCTCAAAT GGGTGAAAGA 12840

AAACGTAAAA GAAGGATTCG ACCATAACCG TGAAGACCTT GTTTTAAGCC GTGAAGAAAA 12900

AGATAGACAA TGGGAATTTG TTATTAAGAT GTTCATGATT GGACGTGACT TAATGGTTGG 12960

TAACCCAAGA CTTGCTGAAC TTGGTTTTGA GGAAGAAGCG GTTGGTCACC ATGCTTTAGT 13020

AGCTGGTTTC CAAGGTCAAC GTCAGTGGAC AGACCATTTT CCAAATGGGG ACTTTATGGA 13080

AACTTTCCTC AATACTCAGT TTGACTGGAA TGGTATTCGA AAACCATTTG TATTTGCGAC 13140

AGAGAATGAT TCACTAAATG GTGTGTCTAT GCTCTTTAAT TATCTATTAA CAAATACTCC 13200

ACAAATCTTT GCTGATGTGC GTACTTATTG GAGCCCAGAG GCTGTTAAAC GTGTAACGGG 13260

ACATACTTTA GAGGGTCGTG CTGCAGCTGG CTTCTTACAT CTAATCAACT CTGGTTCTTG 13320

TACATTGGAT GGTACAGGTC AAGCTACTCG AGATGGCAAA CCTATTATGA AACCATTCTG 13380

GGAGTTGGAA GAAAGTGAAG TGCAGGCTAT GCTTGAAAAT ACAGACTTCC CACCAGCAAA 13440

CCGCGAATAC TTCCGTGGAG GAGGATTCTC AACTCGTTTC TTGACGAAGG GGGATATGCC 13500

AGTAACAATG GTACGTCTCA ATCTTCTAAA AGGGGTTGGT CCAGTGCTAC AAATTGCAGA 13560

AGGTTACACA CTTGAACTTC CTGAAGATGT TCACCATACT TTAGATAATC GTACAGATCC 13620

AGGATGGCCA ACTACTTGGT TTGCTCCACG TTTGACAGGA AAAGGTGCTT TCAAGTCTGT 13680

CTATGACGTC ATGAATAATT GGGGAGCTAA TCACGGAGCC ATAACATATG GACACATTGG 13740

AGCAGACTTG ATTACCTTGG CTTCTATGTT GAGAATTCCT GTCAATATGC ATAATGTACC 13800

TGAGGAAGAT ATCTTTAGAC CTAAAAATTG GTCCTTATTT GGAACAGAAG ATCTAGAATC 13860

AGCAGACTAT CGTGCATGTC AGTTGTTGGG GCCACTACAT AAATAAAACT TGTTTATATA 13920

GGAGGTGAAC TTACGTCCCT CCTATCCTTT TAAAAAGATT TGTTAAACAA TTCACAAATA 13980

ATTGAAAACG AATACAAAAA GTAATATAAT GATGTTAAAT AGATAGCGCG GAGGCGCAGG 14040

AGGAAAATTA TATGGCTATA TTTTATGTTC CGGCAGTCAA CCTTATTGGA AAAGGTGTTG 14100

TAAATGAAGT GGGTCCTTAT ATCAAGGAAC TTGGCTATAA AAAGGCACTT TTGGTGACAG 14160

ATAAGTACAT CGAAGGCAGT GATATTTTAC CTAAGACTTT AAAACCACTG GATACAGAAG 14220

GAATCGAATA T 14231

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16995 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

AGTTCTCTTA ACTTTTTTAG GATGGCATTC TCCGCTCTCA GGTACTCATT TTCTGCTgAA 60

GACGTTCTAA TTCTGTCCTC TCTTCAGGTC TCGTTTTTGG CTTACGTCCC ATTTTAGGTA 120

CTCTCCCTCT TGTTTTCTCA ACAATAGTAT ACCCGTTTTT CCTGTATTGT GCTAGCCAGT 180

TAAGAAGTAT CGTACGACTT GGGAGACCGT ATTCAAGAGA AACTCTATCT TTAGTCCAGC 240

CTTCATGTCA GACTTTATTA CTCATTTCTT GTTTTAAATC AGGAGAATAG TAACGATTTT 300

TTCCTTTTTT GACGAACTCT ATTCCGTAAC GATCAATCAA TTTAATCATG TACCTAATAT 360

TAGAATTGCT TATCCCAAAT TTATTTGAAA GCTTCTCTAA GCTATATCCT TGTTTTCTAA 420

GTTCATAGAT CTGAACTTTA TCATCATAAG TTAGTTTCAT AATAAAAACA CCCCAAAAGT 480

TAGATTTTTT CTGTCTAACT TTTGGGGTGT AGTTCATGTA CACCTGATAT GATGCGTTTT 540

ATAATTTTTA AGCCTTTTTG CCCAGCCTCG TCAAAAGTAA TGTTTTGACA CAAAATCTGT 600

GACAAAACTT TAGTTTTAAA GGTTTTTAAC TTTGTATATA CTAGTTTTAA GAAAAGGAGG 660

ATGATCTAAT GGAAGAAAAA GTATCATTGA AAGTCAGGGT TCAAAAACTA GGGACATCGC 720

TTTCAAATAT GGTTATGCCC AATATTGGAG CATTTATTGC TTGGGGAGTA TTGACTGCCC 780

TCTTTATCGC TGATGGCTAT CTGCCAAATG AACAGTTAGC TACTGTTGTT GGTCCTATGT 840

TAACGTATTT ATTGCCAATC CTGATTGGTT ACACAGGTGG ATATATGATC CATGGCCAAC 900

GTGGTGCCGT TGTAGGAGCT ATTGCTACTG TTGGTGCAAT CACAGGTTCT AGTGTTCCTA 960

TGTTTATCGG AGCTATGGTA ATGGGCCCAC TGGGAGGATG GACTATCAAG AAATTTGATG 1020

AGAAGTTCCA GGAAAAAATT CGTCCCGGAT TTGAAATGTT AGTTAATAAC TTCTCAGCTG 1080

GTCTCGTTGG TTTTGCATTA TTGCTTTTGG CTTTCTACGC AATCGGTCCA GTCGTATCGA 1140

CTCTTACTGG AGCTGTTGGG AATGGTGTTG AGGCTATTGT CAATGCTCGC CTCCTTCCTA 1200

TGGCTAATAT TATCATCGAA CCGGCTAAAG TCCTTTTCCT CAATAATGCC CTCAATCATG 1260

GCATTTTTAC TCCTCTGGGA GTAGAACAGG TAGCTCAAGC TGGTAAGTCA ATTCTCTTCC 1320

TATTGGAAGC TAATCCTGGA CCAGGTCTGG GAATTCTATT AGCTTATGCT GTATTCGGTA 1380

AAGGTTCTGC TAAATCTTCT TCTTGGGGGG CAATGGTTAT TCATTTCTTC GGAGGGATTC 1440

ATGAAATTTA CTTTCCTTAT GTTATGATGA AGCCTACTCT ATTTTTAGCT GCTATGGCAG 1500

GAGGTATCTC TGGAACTTTT ACTTTTCAAC TCTTAGACGC TGGTCTTAAA TCTCCAGCTT 1560

CACCAGGTTC TATTATTGCG ATTATAGCTA CGGCGCCAAA AGGTGTTTGG CCCCATCTAA 1620

ATGTTCTTTT AGGTGTTTTA GTGGCAGCAG TTGTTTCTTT CCTTGTAGCA GCCCTTATTC 1680

TTCATGCAGA CAAGTCAACT GAGGATTCGC TCGAAGCTGC TCAGGCGGCT ACCCAAGCAG 1740

CTAAGGCTCA GTCTAAAGGT CAGTTAGTAT CAACTTCTGT TGATGCAGTT GTTTCGACAG 1800

ACTCAGTGGA AAAAATCATT TTCGCCTGCG ATGCTGGTAT GGGAAGCTCT GCTATGGGAG 1860

CTAGTATTCT TCGAGATAAG GTTAAAAAAG CAGGTCTAGA GATTCCAGTA TCTAATCAGG 1920

CAATCTCAAA TTTGCTTGAT ACACCAAAAA CATTAATTGT TACTCAGGAA GAACTGACAC 1980

CAAGAGCTAA AGACAAGAGT CCAAGTGCTA TTCATGTTTC TGTTGATAAT TTCTTAGCGT 2040

CCTCTCGTTA TGATGAAATT GTAGCTTCAT TAACAGGAGC TTCTCCAATA GCAGAAATTG 2100

AAGGAGATAT ACCAACTTCA GCACCAGTAG ATAGTCAGGA AAGTGACCTT AACCATATTG 2160

ATGCTGTAGT AGTTGCTTAT GGTAAAGCAC AGGGAACTGC AACTATGGGC TGTGAAACGA 2220

TTCGGGCTAT TTTTAGAAAC AAGAATATTC GTATTCCAGT TTCTACTGCC AAAATTTCAG 2280

AATTAGGTGA ATTTAATTCT AAAAACATAA TGATTGTAAC AACTATTTCT TTACAGGCAG 2340

AAGTGCAGCA AGCAGCACCG AATTCTCAAT TTCTTATTGT GGATAGTTTA GTAACAACAC 2400

CAGAATATGA CAAAATGGCT GCTAGAATGT ACAAATAGAA CTAGAGGTTT CTAAATTACG 2460

AATGCTATTA ACCAAACGAG AAGAACAATT ATTGAAGGCT TTCCTACATG TAGGGAAGCT 2520

TTCAATGCAA GATATGACTG AAATCTTACA GGTTTCATCT AGAACAATTT ATCGAACTTT 2580

ATCAGATTTG ACAGATAGCA TGGAGCAATA TGGAATCGAA ATAACGAAGC ATGGGAAATA 2640

CTATATTTTG ACTGGAGAGT TGGATGATTT GCCGACAGAA CTTGAAGTGT TAGTTGAGTA 2700

TAGTCCCCAA GAAAGACAAG AGTTGATTAC CTATCGCCTT CTGACTGAGA GTGGTTTTGT 2760

CACCAATGAA GCATTGCAAG AGTGCACGAA AGTCAGTAAT GTAACTATTA TTCAGGATAT 2820

TTCAGATATT GATAAGCGTC TTTTAGACTT TGATCTGAAA ATTGAACGAC AAAAAGGTTA 2880

TCGGATTTCT GGTGATTCAG TTGGTAAGAG AAGATTTTTG GCTATTTTAC TGACAAACTG 2940

TATCTCAGTA GCAGATTTTT CAACCGGTAA TTTTGGGAGC TTTGATATTT TAGAAGCAGA 3000

TAGAACTGGG CTGGCCAGTC AGATTGTTAA TAAGCAACTG TCAGGTTTTC CAGATATGGA 3060

TGCTAGGATG AAGATGTTTT TTGCGATCTT GTTATCTCTT ATAGGTCAGG AGCAAAACAT 3120

TGAAAATTCA CCTAATACTA GTAAGCAGGC TTTGGAAATT TCTCAAAAAA TTTTTCAAGC 3180

TTACTCTAAG CAGACTGCAC AATTTTATAG TATTCAGGAA ATTATCTATT TTGCGAGCAT 3240

CTTGGATGAA TTAATCATTA AACGTCAGGA CAATCCGCTC TTTACGGAGA AATTTGATGG 3300

TGAATTTTTC TACAATATTT CAAATCTGAT TGATACGGTT TCCATGTATA CCAAGATTGA 3360

CTTTTTTAAG GACAAGGTTT TATTCAATTT TCTTTTCCAT CATATTCGGC TCAGTTTAGG 3420

CGTCCCTATC CTTTTTCAGG GTGAAAATTT GCCAGAATCT ATCCAGATTT TAGTTGAAAG 3480

GAATAAATTT CTTTATACAG TCATCAGTCT TTTAGTGAAT GATATTTTTC CGAAATATCT 3540

TCATACAGAG TATGAGTATG GCATGATTGC CCTACATTTT ATCTCTAGCT TAGGCCGTAG 3600

TCCAGAGATT TATCCAGTCC GTGTTTTGCT TTTAACGGAT GAACGTCGGG TCACTAGAGA 3660

TTTATTAGTC AGTAAAATTA AGAGTGTTGC TCCTTTTGTA GAGTTGATAG ATATTCAATC 3720

TCTAGTAGAT TACCACAGTA TTGATCTCAG TCAGTATGAT TATATTTTAT CTACCAAGCC 3780

GCTGACTAAT CAGGAAATCG ATGTAATTTC TAGTTTTCCA ACCGTCAAAG AATTGCTTGA 3840

ATTACAGGAA CGACTTCAGT ATGTACAGGC ACATCGTACA ATTGTCGCGC GTGATGCTAT 3900

CGCTCCAGAG AAAAGTTATG ACTTGCAAGA TTATTTAATA TCTAGTAGTC AGCTTTTGAG 3960

TCAATTCGAG TTGGTTCAAT TGGAGAATAA TCAATCATTT GAGCACACGG TAGAACAAAT 4020

CATCCAATAT CAGAAGAATG TGAGTGACAG AGCTTACCTA ACAAGAAAAT TGTTATCTCA 4080

CTTCCAGAAT AGTCCTATGG CTATTCCTAA TACTGGTCTG GTGCTTTTAC ATAGTCAGTC 4140

TAGCAAAGTA ACAACAAATA GTTTTACTAT GTTTGAACTC AAACTACCTA TCTCCGCATT 4200

GTCAATGAAA CGAGAGGAAG AAGAGGTCAA AAGGTGTCTG CTAATGCTAA TGTCTAAAGA 4260

AGCTAGCGAG GAAGCTAGAG ATTTAATGAC AGCTATTAGT CAGTCGATTA TTGAAAATCA 4320

TCTTTATACA GAGATTTACA AGACGGGAAA TCAATCCATT ATTTATCAGA TGCTAAATAC 4380

TATTTTTAAC GAAAAAATTA AGAAATTGGA GAACTAATAT GAAACTTGAA AAACATTTGA 4440

TTAAGCTTAA TAAACAATTT TCTAACAAGG AGGAAGCTAT TTGTTATTGT GGGCAAGTTC 4500

TTTATGAGGG TGGATATGTT AATGAAGACT ATATTGAAGC CATGATTGAG CGAGATAAAG 4560

AGCTATCTGT TTACATGGGT AACTTTATCG CCATACCGCA TGGAACAGAT GCAGCAAAAA 4620

ATGATGTCCT CAAGTCTGGT ATTACAGTCG TTCAAGTCCC TAGAGGGGTT GATTTTGGGA 4680

ATGTATCTAA CCCTCAAGTG GCAACGGTTC TTTTTGGTAT TGCTGGTATT GGTAATGAAC 4740

ACTTAGAAAT TATTCAGAAA ATTTCTATCT TCTGTGCAGA TGTAGATAAT GTTCTTAAAC 4800

TAGCAGATGC TCAGTCAAAA GAGGAAGTAT TGCGCTTATT TGATGCTGTT GAATAATTGA 4860

ATTTAGTCAT TTGTCATCTA GTATATATGT CCCTCAAATA GGAAAAGGAG AAATTGAATG 4920

AAACATTCTG TTCATTTTGG TGCCGGTAAT ATCGGTCGTG GTTTTATAGG TGAAATTCTA 4980

TTTAAAAATG GTTTCCATAT TGATTTTGTG GATGTCAATA ATCAGATAAT TCATGCTCTG 5040

AATGAAAAGG GCAAGTATGA AATTGAAATT GCACAGAAAG GACAGTCTCG TATAGAAGTA 5100

ACTAATGTGG CTGGCATTAA TAGCAAAGAA CATCCTGAGC AAGTCATTGA AGCGATTCAA 5160

AAGACGGATA TTATTACTAC TGCAATCGGA CCTAATATAC TCCCTTTTAT CGCCGAACTT 5220

CTAGCCAAAG GAATCGAAGC TCGCCGAGTT GCAGGAAATA CACAGGCATT GGATGTTATG 5280

GCCTGTGAAA ATATGATTGG CGGGTCTCAA TTTCTTTATC AAGAAGTCAA GAAATATTTA 5340

AGTCCGGAAG GTTTGACATT TGCTGATAAC TACATAGGTT TTCCAAATGC TGCAGTAGAC 5400

AGGATTGTTC CAGCACAAAG TCACGAAGAT TCCCTTTTTG TTGTGGTCGA GCCCTTTAAT 5460

GAATGGGTCG TGGAAACCAA GCGTCTTAAA AATCCAGATT TACGTCTAAA AGATGTGCAT 5520

TATGAAGAAG ATTTAGAACC CTTTATTGAG CGAAAACTTT TTTCAGTCAA TTCTGGACAT 5580

GCAACTTCAG CTTACATTGG TGCGCATTAT GGTGCCAAGA CAATTTTGGA AGCTCTTCAA 5640

AATCCTAATA TTAAATCTCG GATTGAATCT GTATTAGCTG AAATTCGGAG TCTCTTGATT 5700

GCCAAATGGA ACTTTGATAA AAAAGAATTG GAGAATTATC ACAAAGTCAT TATAGAACGA 5760

CTTGAAAACC CTTTCATAGT GGACGAGGTT AGTCGCGTAG CTCGTACTCC AATCCGAAAA 5820

TTAGGCTATA ATGAACGATT CATCCGGCCG ATACGTGAAT TGAAAGAACT CAGTTTGTCA 5880

TATAAAAACC TACTTAAAAC AGTTGGCTAT GTCTTTGACT ATCGCGATGT AAATGATGAA 5940

GAAAGTATTC GATTAGGTGA ATTGTTGGCT AAACAATCAG TCAAAGATGT TGTTATACAA 6000

GTTACAGGTT TAGACGACCA AGAATTGATT GAGCAAATTG TAGAGTATAT TTAATCTTTT 6060

TCGAAAATCT CTTCAAATCA GGTTAGCATC GCTTTGTCTT AGGCATATGT TGTTCTATCT 6120

ACAACCTCAA AGCAGTGCTT TGAGCTGACT CCGTCAGTCT TATCTGCAAT CTCAAAACAC 6180

TGTTTGAGTT ATCTGCGGTA ATCTTTCTAG CTTGTCTTTG ATTTTTGTTG TTATTTATAA 6240

GGTAAAAGAA GCTGGACAAA AAGTCTTCAA AATCGGGAAA AGGCAGCCTA TCGGGTGTTC 6300

AAAAATCTTG ATAGGATGTC CTTTATTATG GAAAGCCTTA TTGGATTTTC TCCTCAGATT 6360

GAGTTTTTGA TCAGCTTTAT GAGATAGGTC TTGCTAGAGA TGTAGCCCAT CATGTTATTT 6420

TTATGGACAG TGGGAAAATT GTTGAAAAAA ATAATGCCCA TCAATTCTTT AGTCGTCCAA 6480

GAGAAGAACG AACCAAGCAA TTTTGGAACG AATTCTTTCG AATGCGATCT ATATAGTAAA 6540

ATGAAACAAG AACAGGACAA ATCGATCAGG ACAGTCAAAT CGATTTCTAA AAATGTTTTA 6600

GAAGTAGAGG TGTACTATTC TAGTTTCAAT CTACTATATA ACTGAAAAAT TAGATAAATT 6660

AGTTTTGGAA AATGACTAAC CAAAAGATAT CCAAAGTAGT CTAAAATTGT CTATACTTTA 6720

TGAGTGTTTT AGTTAGGAAA AAGGCTTGTT GTCTATAATT GTCTGCATTA GTCTAGATTT 6780

TATTTATAGA AAATGTTATA ATAGACTGTA TTTAAAAAAT TTTAAGGAGA AATGACAGAA 6840

TGTCTGTATC ATTTGAAAAC AAAGAAACAA ACCGTGGTGT CTTGACTTTC ACTATCTCTC 6900

AAGACCAAAT CAAACCAGAA TTGGACCGTG TCTTCAAGTC AGTGAAGAAA TCTCTTAATG 6960

TTCCAGGTTT CCGTAAAGGT CACCTTCCAC GCCCTATCTT CGACCAAAAA TTTGGTGAAG 7020

AAGCTCTTTA TCAAGATGCA ATGAACGCAC TTTTGCCAAA CGCTTATGAA GCAGCTGTAA 7080

AAGAAGCTGG TCTTGAAGTG GTTGCCCAAC CAAAAATTGA CGTAACTTCA ATGGAAAAAG 7140

GTCAAGACTG GGTTATCACT GCTGAAGTCG TTACAAAACC TGAAGTAAAA TTGGGTGACT 7200

ACAAAAACCT TGAAGTATCA GTTGATGTAG AAAAAGAAGT AACTGACGCT GATGTCGAAG 7260

AGCGTATCGA ACGCGAACGC AACAACCTGG CTGAATTGGT TATCAAGGAA GCTGCTGCTG 7320

AAAACGGCGA CACTGTTGTG ATCGACTTCG TTGGTTCTAT CGACGGTGTT GAATTTGACG 7380

GTGGAAAAGG TGAAAACTTC TCACTTGGAC TTGGTTCAGG TCAATTCATC CCTGGTTTCG 7440

AAGACCAATT GGTAGGTCAC TCAGCTGGCG AAACCGTTGA TGTTATCGTA ACATTCCCAG 7500

AAGACTACCA AGCAGAAGAC CTTGCAGGTA AAGAAGCTAA ATTCGTGACA ACTATCCACG 7560

AAGTAAAAGC TAAAGAAGTT CCGGCTCTTG ACGATGAACT TGCAAAAGAC ATTGATGAAG 7620

AAGTTGAAAC ACTTGCTGAC TTGAAAGAAA AATACAGCAA AGAATTGGCT GCTGCTAAAG 7680

AAGAAGCTTA CAAAGATGCA GTTGAAGGTG CAGCAATTGA TACAGCTGTA GAAAATGCTG 7740

AAATCGTAGA ACTTCCAGAA GAAATGATCC ATGAAGAAGT TCACCGTTCA GTAAATGAAT 7800

TCCTTGGGAA TTTGCAACGT CAAGGGATCA ACCCTGACAT GTACTTCCAA ATCACTGGAA 7860

CTACTCAAGA AGACCTTCAC AACCAATACC AAGCAGAAGC TGAGTCACGT ACTAAGACTA 7920

ACCTTGTTAT CGAAGCAGTT GCCAAAGCTG AAGGATTTGA TGCTTCAGAA GAAGAAATCC 7980

AAAAAGAAGT TGAGCAATTG GCAGCAGACT ACAACATGGA AGTTGCACAA GTTCAAAACT 8040

TGCTTTCAGC TGACATGTTG AAACATGATA TCACTATCAA AAAAGCTGTT GAATTGATCA 8100

CAAGCACAGC AACAGTAAAA TAATCTTAAT AAACAGAAAA CCCACCTGAA TTGGTGGGTT 8160

TTCTGATGCA CTATTTTCCA AAAATCTCTT TGAGGTCTGT GTCTGTAATC CCAATCATGG 8220

CTGGGATGCG GTCCCAGTTT TCTTCGGTTA GGATGTAGGA TTGTTCAGAG GCACTTGATG 8280

TGACTGTTTC AGAGACAGCT TGTTGCTTTT CTTCAACATT CTCCAGTAGA TCACTGAAGC 8340

GTTCAATCAG ATAGGTTTTT CGGGCAGTTC CGATGTGTTG GGTAGCATAG TCGAAGGCTT 8400

GTAATTCGCC TAGTAAGATG AGTTTGCTTT TGGCACGTGT AATGGCTGTG TAGATGAGAT 8460

TTCGCTCCAG CATACGTCGG CTAGCACTAG TAATCGGTAG GATGACAACT GGGAACTCAC 8520

TTCCCTGAGA CTTATGAATA CTCATGGCAT AGGCCAAGCG AATCTTGTAC CATTCGTTAC 8580

GGGGGTAAGA GACTTCATTA CCATCAAAAT CAATGACAAT CTCGTCTTGT TTCGATTCGG 8640

TGTATTTACC AGGAATCAGG TCTGTGATAG CTCCTAAATC CCCATTAAAG ACATTGATTT 8700

CAGCATCGTT AACCAAATGA ATGACCCTGT CTCTCTTACG ATAGTGACAC TGAGGAGCTT 8760

CAAAACTGAG TTGATCTTTT TGTGGGGGAT TGAGCAGGTC TTGCATGAGC TGATTGATAG 8820

CATCAATCCC TGCCGTCCCT CGGTACATAG GAGCCAGAAC TTGGATATCA CGGGCGGGAA 8880

TACCATTTCT GAGGGCGGCA CCTAAGATTT TTTCAATGGT GGCAGGAATA TGGCCACTAG 8940

CAATTTCAAA GTAGGAACGG TCAGCTTTTT TTTGGGTGAA ATCAGCTGGC AAGATGCCCT 9000

GTCGAATCTG ACTAGCTAGG GTGACGATGG TTGATTCTTT GCTTTGTCGA TAAATTTTTT 9060

CCAAGCGAGT CTGAGGAATC AAAGGAATAT GAAGTAGATC CGCTAGAACC TGTCCAGGAC 9120

TGACAGAAGG TAGCTGATCA CTGTCACCTA CGATGAGGAT CTTACTGTTA GAAGAGATAT 9180

TGGAGAAGAG TTGATTGGCC AGCCAAGTAT CTACCATAGA GAATTCATCC ACGATGATAA 9240

AGTCAGCATC TAGGTAATCT TCCAGATGAC TGGTATCATC GTCACCTGTC ATTCCCAAGT 9300

GGCGATGTAT GGTCGCGCTA GGCAAACCTG TCAATTCATT CATGCGACGA GCAGCTCGAC 9360

CAGTTGGAGC AGCAAGAAGA ATGGGCAGAT TGCTTTTCTT CCTGAAGTCA AGTCCTTCTA 9420

AAAGGGCATA AACAGCAATG ATTCCATTGA TAACAGTTGT CTTACCAGTA CCAGGCCCAC 9480

CTGTCAGGAT AAAGACCTTA TTCTGGATAG CATCACAGAT AGCCTGTTTT TGAATGTTAT 9540

CATACTCAAT TCCCAGTTCT TGCTCGACAG TAGTGATATG TTTTTGAATG GTTTCTAAAT 9600

CATGACTCTT CTGTTTTCCT TTTTCAAGGA TACGAACCAA GTGACTGCGG ATGCCTTCCT 9660

CAGCGAAAAA GAGGCTGTTG TCAAAGATCT TGGTATCAAT CTGCTGAACC TTGTCTTCTT 9720

CGATCAGGTA GGAGAGCTCT TGGGCAACTT GGCTGGGGTC TAGTTCCACG GGACGGGAAG 9780

ACTCAAGGAG AGTAAGGGTT TGTTCCAGCA AATCCCGTGC TTCAACATAG GTGTCCCCTG 9840

TTTCCATACA GGCCTGAAAA AGACTGTGAA CTAGACCGGC GCGGAAGCGT TCAGGAGCCT 9900

GACTTTCGAT GCCTAGTTCC TCAGCTAGTT GGTCAGCAAT GGTAAAGCCC AAACCCTTGA 9960

TATCCTCAAC CAGCTGGTAG GGATAATTTT CAACCACATC AAGGGTTTCT TCCTTGTAAA 10020

AGTCTTGAAT CTGAAAGGCT AGTTTGTTGG GAATGCCGTA GTTGGCTAGT TTGGCCAAAA 10080

TCATCTCCGT TCCGTAGTTG AGACGGAGAG TGGAGACGAA AGCCTCGCGA TTTTTGGCAG 10140

AGAGTCCTGC GATGCCTTCT AACTTTTCTG GGTGTTGCAA AATTTCGTCA ATGGTATTTT 10200

CGCCATAGGT ATCCACGATT TTCTGAGCTG TCTTGAGACC AATCCCCTTG AAATGGCTAC 10260

TTGAAAAGTA CTTGACCAAG CCCTTACTAG TTGGTTTTGC GCGATCATAA CGACTGATTT 10320

GCAGTTGTTC TCCATACTTG GAGTGCTGGA CAATTTGCCC CCAAAAAGTA TAGTCTTCGC 10380

CCTCAATTAC ATCAGCCATG GTTCCTGTGA CAATGATTTC AAAATCATCA AAATCCTCTG 10440

CGTCCGTATC GTCGATTTCT AGGAGGAGGA TGCGATAAAA ATTGCTGGGA TTTTCAAAAA 10500

TAATCCGTTC AATAGTTCCT GAAAAATAAA CTTCCATAAA ATTCCTTTGC ATGAATAGGT 10560

GAGAGTTGGG ATTGTTTTTA TTTTATACTC TTCGAAAATA TCTTCAAACC ACGTCAGCTT 10620

CCATCTGCAA CCTCAAAACA GTATTTTGAG CTGACTTCGT CAGTTCTATC CACAACCTCA 10680

AAACACTGTT TTAAGCAGCC TACGGCTAGC TTCCTAGTTT GTTCTTTGAT TTTCATTGAG 10740

TATTTGTAAA TAAACAATCA CTTCTCACGA TAGAAGAAGA GGCTGAGATT GGTGATTCTC 10800

TGCCTCTTAG GTTTCTTAAA ATGTTCCGAT ACGGGTGATT GGCCATAAGC GGAATTTAGC 10860

TTCCCCTGTG ATATCTTTTG CTTTGAAGGT ACCTACGTGG CGGCTGTCGC TCGAAACCAA 10920

GCGGTCATCT CCGAGGAGAA GGTATTCTCC TTCTGGAACA GTAAAGCTAA AGTTGGTGTT 10980

GTAGTTGACA TCAACTGTGA AGGCTTGAGC TTTTTGAGCG ATACTTCTAA AGAAAGTTCC 11040

TTTATTTCCT TCAAAGCCCT TGCCTGAGTA AGTGCTTTGG AGTTTGTCAT CCTTGAAGCG 11100

TTTGATATAG TCTGCTAGAT AAGGCTCGTC CGTTTCTTTG TCATTGATGT AGAGTTTATC 11160

ATTTTCGTAA CGAATGGTGT CGCCAGGCAT TCCAATCACG CGCTTGACGA TGTCCTTATT 11220

GCCATCTTCC TCATGGGCCA CCACGATATC AAAACGGTCA ATAGGAAGGT GTTTTACAAC 11280

GAAGAGAATT TCGCCATCCG CTAGGGTCGG ATCCATGGAA TGTCCTTCTA CGCGAACATT 11340

GCTCCAAAAA AAGATACGAC TTAAAGCTAG TAATGACAGA ATTAGGAGGA ACAATCCCCA 11400

CTCTTTTAAG AAATTTTTAA ATGAATTCAT AACTTACCTT TCTAAGCGTT TTTTCGCTTT 11460

TTCAGTGTTT TTAAAGTGCA ATTTGGCGCA GAAGCTGAGT CCCTGCATAC CATAGGCTTG 11520

CAAAATCTGG CTAGCCACCT TGTCAGAAGC CGTTCCAGCT CCACTTGGGA GCTGATAACC 11580

CAGTTCTCGT CCCAAATTTT CAAGATTTTC CAGAAAGAGA TCACGCGCAA TGACAGAAGA 11640

AACTGCGACA GACAAGTATT TGCCCTCAGC CTTTTCTTCT AAGCTGATAG GATTGCTGAA 11700

ACGATTGGCC TCTTGTGCCA AGTACTTGTC ATAATTTTTA GCACTGGTAA AGGCATCAAT 11760

CACAATTTTC TCAGGCTGAA CACCTTTTTG AAGGAGGAGA TAGATAGCCT GATTATGGAG 11820

GGCAACCTTA ACCGAAACAG CGTTGTAGCG GTCTCCGATG ACCTCGTTGT ACTTGCTGGG 11880

TGAGAGAAGG AGTGCCTGGT GCTGAATTTT TTCCTTGAGA ATAGGAGTAA TCTGACGGAT 11940

CTTTTGGTCG GTCAGAGTCT TAGAATCCCC CACACCGAGT TTTCGTAAAA AGTCGTGCTG 12000

GTCAGGTGTG ACAAAGGCAG CCACAACTGC AAGCCCACCA AAGTAGGAAC CATTTCCCAC 12060

CTCATCTGTC CCAATTAAAG GAAGATTTTG TCCGCTGGTT TGCTCTACAG CTTGATAGCC 12120

AAAGAAACTG GCGTATTTTT CAGCCCCTTC ACCCTGAAGC AAGATTTTTC CAGAAGTATA 12180

GATAGAAACC GTTGCTTGAG GTAGTTTCAA AAAGTAGCGG ATATAGGGAT TCTTGCTAGG 12240

AGCCAGACTG GTTTGATAGT GTTCAAGAAA AGCCTGAATA TCCTTTTCGC TTGGTGTGAG 12300

TGTGATACTT GCCATAGTTT CTATTGTACC ACAAAAGCAG TAAAATTTGT AAAAACTGAC 12360

AAAATTAGCG AATTTTGGTA TAATATCGTG AGGTGAATTT TATGGCAAAT CTAAATCGAT 12420

TCAAATTTAC ATTCGGGAAA AAATCGTTAA CCTTGACAAG CGAACATGAC AACCTTTTTA 12480

TGGAGGAAAT CGCTAAGGTT GCGACAGAAA AATACCAAGC AATTAAAGAA CAAATGCCTA 12540

GCGCAGATGA TGAAACAATC GCTCTTTTGT TGGCAGTCAA CTGTTTATCA ACTCAGCTCA 12600

GCCGTGAGAT TGAATTTGAC GATAAGGAGC AAGAGCTAGA AGAACTCCGT CACAAGCTTG 12660

TGACTTGTAA GCAAGAACAG AGCAAGATTG AGGATTCCTT ATGATTTCAT TCCTTCTTCT 12720

ATTGGTCTTG GTTTGGGGAT TTTATATCGG CTATCGGAGA GGCCTGCTCT TACAGGTTTA 12780

TTACCTGATT TCAGCCATGG CATCGGCTTT TATGGCTGGC CAGTTTTATA AGGGGCTTGG 12840

AGAGCAATTC CATTTATTGC TCCCTTATGC AAATTCGCAG GAAGGTCAGG GGACTTTCTT 12900

TTTCCCATCG GATCAACTCT TTCAGCTGGA TAAGGTCTTT TATGCAGGTA TCGGCTACTT 12960

GCTTGTATTT GGGATTGTCT ATAGCATTGG TCGTTTACTT GGTCTTCTCT TACACTTGAT 13020

TCCTAGCAAA AAACTGGGTG GTAAGTTGTT CCAAGTTTCA GCAGGTATCT TGTCCATGTT 13080

GGTGACCTTA TTTGTCTTGC AAATGGCCTT GACAATCTTG GCGACCATCC CCATGGCAGT 13140

TATACAAAAT CCTCTTGAAA AGAGTATCGT CGCAAAACAC ATCATCCAGA GCATACCGGT 13200

AACAACCAGT TGGCTCAAAC AAATCTGGGT GACAAATTTA ATCGGATAAA AAGGGCAGGA 13260

GTTTTCCTAG CCCTTTGTTT ACAGATTTGA CTCGAATCTA TCAGAATGTA AAAAGCTACC 13320

ACACCTAGAC ATTCAAAGAC AAGGAAATAA AGATGAATAA GAAAATATTA GAAACATTAG 13380

AGTTCGATAA GGTCAAGGCC TTGTTTGAGC CTCATTTGTT GACCGAGCAG GGCTTGGAGC 13440

AATTGAGACA ACTGGCTCCG ACTGCCAAAG CAGATAAAAT CAAACAGGCT TTTGCTGAGA 13500

TGAAGGAAAT GCAGGCTCTT TTCGTCGAGC AACCGCATTT TACTATTCTC TCAACTAAGG 13560

AAATTGCAGG AGTCTGCAAG AGGTTGGAGA TGGGAGCGGA TCTCAATATC GAGGAGTTCC 13620

TACTCTTGAA ACGCGTGCTT CTTGCCAGCC GAGAACTTCA AAATTTTTAC ACCAATCTGG 13680

AAAATGTCAG CTTGGAAGAA TTAGCCCTTT GGTTTGAGAA ATTACATGAT TTTCCGCAAT 13740

TACAAGGAAA TCTTCAGGCC TTTAATGATG CGGGTTTCAT TGAAAATTTT GCCAGTGAAG 13800

AATTGGCGCG AATCCGTCGA AAAATACATG ATAGCGAGAG TCAGGTACGC GATGTTTTAC 13860

AAGACTTGCT CAAGCAAAAA GCGCAGCTGT TGACGGAAGG AATTGTTGCT AGCAGAAATG 13920

GCCGTCAGGT TTTACCAGTC AAAAACACCT ACCGCAATAA GATTGCAGGT GTCGTTCATG 13980

ATATTTCTGC TAGTGGAAAC ACCGTCTATA TCGAACCCCG TGAGGTAGTC AAACTGAGCG 14040

AAGAAATTGC TAGTCTGCGA GCAGATGAGC GCTATGAAAT GCTTCGCATT CTCCAAGAAA 14100

TTTCTGAGCG TGTCCGCCCT CATGCGGCTG AGATTGCTAA TGACGCTTGG ATTATCGGTC 14160

ATCTGGACTT GATTCGTGCC AAGGTTCGAT TTATCCAAGA AAGACAAGCA GTCGTGCCTC 14220

AGCTGTCAGA AAATCAAGAG ATTCAACTGC TCCATGTCTG CCATCCTTTG GTCAAAAATG 14280

CCGTCGCAAA TGATGTCTAT TTTGGTCAAG ATTTAACAGC TATTGTCATT ACAGGTCCCA 14340

ATACAGGTGG GAAGACCATC ATGCTCAAAA CTCTGGGCTT GACACAGGTC ATGGCCCAGT 14400

CAGGATTGCC GATTTTAGCA GACAAGGGAA GTCGTGTTGG TATTTTTGAA GAAATCTTTG 14460

CTGATATTGG AGATGAGCAG TCTATTGAGC AGAGCTTGTC TACCTTCTCT AGTCATATGA 14520

CCAATATCGT GGATATTCTT GGCAAGGTCA ACCAACATTC ACTCTTACTT TTGGATGAGT 14580

TGGGGGCTGG TACTGATCCC CAAGAGGGAG CAGCCCTTGC CATGGCTATT CTGGAGGACC 14640

TTCGCCTGCG TCAAATCAAG ACCATGGCGA CGACCCACTA TCCAGAACTC AAGGCCTACG 14700

GTATTGAGAC AGCCTTTGTG CAAAATGCCA GTATGGAGTT TGATACTGCA ACTCTTCGCC 14760

CGACCTATCG CTTTATGCAG GGTGTTCCTG GCCGAAGTAA TGCCTTTGAA ATTGCCAAAC 14820

GTCTAGGCCT ATCTGAAGTT ATCGTAGGAG ATGCCAGTCA GCAGATCGAT CAGGACAATG 14880

ACGTCAATCG TATCATTGAG CAATTAGAAG AGCAGACGCT GGAAAGCCGC AAACGTTTGG 14940

ACAATATCCG TGAGGTGGAG CAAGAAAATC TCAAGATGAA CCGTGCGCTA AAAAAACTCT 15000

ACAACGAGCT TAATCGTGAA AAGGAAACCG AGCTTAACAA GGCGCGTGAA CAGGCTGCTG 15060

AGATTGTGGA TATGGCCCTA AGTGAAAGTG ACCAGATTCT CAAAAATCTC CACAGTAAAT 15120

CCCAACTCAA GCCCCACGAA ATCATTGAAG CCAAGGCCAA GTTGAAAAAA TTGGCTCCTG 15180

AAAAAGTGGA CTTGTCTAAA AATAAGGTCC TTCAAAAGGC CAAGAAAAAA CGAGCTCCAA 15240

AGGTGGGAGA TGATATCGTG GTTCTCAGTT ATGGTCAGCG TGGTACCTTG ACCAGTCAAC 15300

TCAAGGACGG TCGCTGGGAA GCCCAAGTTG GCTTGATTAA GATGACCTTG GAAGAGAAAG 15360

AGTTTGATCT TGTTCAAGCC CAGCAAGAAA AACCAGTCAA GAAGAAACAG GTCAATGTTG 15420

TGAAACGAAC TTCTGGGCGA GGACCTCAAG CTAGACTGGA TCTTCGAGGC AAGCGCTATG 15480

AAGAAGCCAT GAATGAGCTA GATACCTTCA TCGACCAAGC CTTGCTTAAC AATATGGCTC 15540

AAGTTGATAT CATCCATGGT ATCGGAACAG GAGTCATCCG TGAAGGAGTT ACCAAATACT 15600

TGCAAAGAAA CAAACATGTC AAGAGTTTCG GCTATGCCCC ACAAAATGCT GGAGGCAGTG 15660

GTGCGACTAT TGTCACTTTT AAAGGATAGC AGTATTCTGG ACTTTATAAA GTAAAAACTG 15720

TTGAACTAAT TTTTACTAAT AAACACATTG ACAAAAGCCA ACATTTTTTG TAAAATTAGA 15780

ATCAATTAAA TACCAACACC GAATGAAGTT TAATAGAAGT GGGGAATCGT TTGATTTTCC 15840

ATGACTGTAA ATGGACGGAA CTCTGGAGAG ACCGTAAAGG CACCGAAGGG CAAGGCAGGC 15900

AACTGCTCAA ACTCTCAGGT AAAAGGACAG AGCTAGGATA GACCGCTTTT TAGCATTTAT 15960

CTAAGCATTC CAGAGTACAT GTATCTTGCA TGTGCTCTTT CTTTTGGGGT TGAAACGATA 16020

GGAGAAGGAA ATGTTAGAAT TGCTTAAATC AATCGATGCT TTTGCTTGGG GACCGCCCCT 16080

CTTGATTTTA TTGGTCGGAA CAGGGATTTA CCTAACTATT CGGCTAGGAC TCTTGCAGGT 16140

TTTGCGTCTA CCCAAGCCCT TTCAGCTTAT TTTTATCCAG GATAAGGGAC ATGGTGATGT 16200

ATCCAGTTTT GCAGCTCTGT GTACAGCCTT GGCATCAACT GTTGGAACAG GAAATATCAT 16260

AGGAGTTGCG ACGGCTATCA AGGTTGGTGG ACCAGGAGCT CTATTTTGGA TGTGGATGGC 16320

GGCTTTCTTT GGAATGGCTA CCAAGTATGC GGAAGGACTC TTGGCCATCA AATACCGCAC 16380

CAAGGACGAC CATGGTGCAG TAGCGGGAGG TCCCATGCAT TATATCCTTC TAGGGATGGG 16440

AGAAAAGTGG CGACCACTTG CTGTTTTGTT TGCAGTAGCA GGAGTATTGG TTGCTCTCTT 16500

GGGAATCGGA ACCTTCACCC AAGTCAACTC GATTGCAGAA TCTATCCAAA ATACAACGAC 16560

GATTTCGCCA GCCATCACAG CTCTCGTCTT GTCTGTCTTT GTAGCGATTG CAGTCTTTGG 16620

TGGACTCAAG TCTATTTCTA AGGTTTCAAC TACTGTTGTT CCTTTTATGG CCATCATTTA 16680

TATCTTAGGA ACTCTTACAG TTATTTTCTT TAATATCGGA AAAATCCCTG GCACAATCGC 16740

TTTAGTCTTT ACCTCAGCTT TTAGTCCCCT TGCTGCGGTA GGTGGATTTG CTGGTGCTAG 16800

CGTTCGGATG GCTATTCAAA ATGGTGTGGC GCGTGGTGTG TTCTCAAACG AATCTGGTCT 16860

GGGTTCTGCT CCTATTGCAG CTGCAGCTGC CAAGACAAAT GAACCAGTAG AGCAAGGTTT 16920

GATTTCCATG ACAGGAACCT TTATTGATAC CCTCATCATT TGTACTCTAA CTGGTTTGAC 16980

CATCTTGGTA ACTGG 16995

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28473 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

CCGGGGCTTT TGTAGTATAA TAGAGATACG TTTTGAAAGT AGGAGGTATC TATGGACTTA 60

ACTAAGCGCT TTAATAAACA GTTAGATAAA ATTCAAGTTT CGTTGATTCG TCAGTTTGAC 120

CAGGCTATTT CGGAGATTCC TGGGGTCTTG CGTTTGACCT TGGGGGAACC TGATTTTACA 180

ACGCCAGACC ATGTCAAGGA GGCGGGCAAG CGAGCGATTG ATCAGAACCA ATCCTACTAT 240

ACAGGGATGA GTGGTCTGCT GACTCTACGT CAGGCAGCCA GTGACTTTGT TAAGGAAAAG 300

TACCAACTGG ACTATGCTCC TGAAAATGAA ATCTTGGTTA CAATTGGGGC GACAGAGGCT 360

TTATCTGCGA CTTTGACGGC TATTTTGGAA GAGGGAGACA AGGTACTTTT GCCAGCTCCT 420

GCTTATCCAG GCTATGAACC GATTGTTAAC TTAGTTGGGG CAGAAATTGT TGAGATTGAT 480

ACGACTGAAA ATGGTTTTGT CTTGACTCCT GAGATGTTGG AGAAGGCCAT TTTGGAGCAG 540

GGTGATAAGC TCAAGGCGGT TATTCTCAAC TATCCAGCCA ATCCGACAGG AATTACCTAC 600

AGTCGAGAGC AGTTAGAGGC CTTGGCAGCT GTTTTACGCA AGTACGAAAT TTTTGTTGTC 660

TGTGATGAGG TTTACTCAGA ATTGACCTAC ACAGGCGAAG CCATGTGTCT CTAGGAACGA 720

TGTTGAGAGA CCAGGCTATT ATTATCAATG GTTTGTCTAA ATCGCATGCC ATGACAGGTT 780

GGCGTTTGGG GCTGATTTTC GCTCCTGCGA CCTTCACAGC CCAGTTAATC AAGAGTCACC 840

AGTACTTGGT CACTGCCGCA AATACCATGG CGCAACATGC TGCGGTAGAA GCCTTGACGG 900

CTGGTAAAAA CGATGCGGAC CCATGAAGAA GGAATATATC CAACGTCGGG ACTATATCAT 960

CGAAAAAATG ACTGCTCTTG GTTTTGAGAT TATCAAACCA GACGGTGCCT TCTATATTTT 1020

TGCTAAAATT CCAGCGGGCT ACAATCAAGA CTCCTTTGCT TTTCTGAAGG ATTTTGCTCA 1080

GAAGAAGGCC GTTGCCTTTA TCCCTGGTGC AGCCTTTGGA CGTTACGGGG AAGGCTACGT 1140

CCGCCTATCT TATGCAGCCA GCATGGAGAC TATCAAAGAA GCCATGAAAC GACTTGAGGA 1200

GTACATGAGA GAAGCATGAT TCAGTCTATC ACGAGTCAAG GCTTGGTGCT TTACAATCGC 1260

AATTTTCGTG AGGATGACAA GCTCGTCAAA ATTTTTACAG AGCAGGTTGG CAAACGCATG 1320

TTTTTTGTCA AACACGCTGG TCAGTCTAAG CTGGCGCCTG TTATTCAGCC CTTGGTGCTG 1380

GCACGATTTC TCTTGCGAAT CAATGATGAC GGACTCAGTT ACATCGAAGA CTATCATGAG 1440

GTCATGACTT TTCCCAAGAT TAATAGTGAC CTCTTTGTCA TGGCCTATGC GACCTATGTG 1500

GCAGCTCTTG CAGATGCTAG TTTGCAGGAC AATCAGCAGG ATGCTCCCTT GTTTGCTTTT 1560

TTGCAAAAGA CTTTGGAGTT GATGGAAGCA GGCTTGGATT ATCAGGTTTT GACCAATATT 1620

TTTGAAATTC AAATTTTGAC TCGATTTGGA ATCAGCCTCA ATTTTAATGA GTGTGTCTTC 1680

TGCCATCGGG TTGGTCAGGC TTTTGACTTT TCTTTCAAAT ATGGAGCCTG CCTCTGTCCA 1740

GAGCATTATC ATGAGGATAA GAGACGTTGT CATCTCAATC CCAATATCCC CTATCTGCTC 1800

AATCAATTTC AAGCTATTGA TTTTGAGACT TTGGAGACCA TTTCGCTCAA GCCTGGAATC 1860

AAGCAAGAGC TACGCCAATT TATGGATCAA TTATATGAAG AGTACGTTGG GATTCACCTA 1920

AAATCAAAGA AATTTATTGA TTCCCTAGCA GACTGGGGAC AATTACTAAA AGAGGAAAAG 1980

AAATGAAAAA AATCGCAGTA GATGCCATGG GGGGCGATTA CGCACCTCAG GCCATTGTTG 2040

AGGGTGTCAA TCAAGCCCTA TCTGACTTTT CAGATATCGA GGTTCAACTT TACGGAGATG 2100

AAGCTAAAAT CAAGCAATAT CTGACAGCGA CAGAGCGCGT CAGCATTATC CATACGGATG 2160

AGAAGATTGA TTCGGATGAT GAACCTACGA GAGCTATTCG GAATAAGAAA AATGCCAGTA 2220

TGGTATTGGC AGCCAAGGCT GTCAAAGATG GTGAAGCAGA CGCTGTCCTT TCGGCTGGGA 2280

ATACAGGTGC CTTGTTGGCA GCAGGATTCT TCATCGTGGG TCGTATCAAG AATATCGACC 2340

GTCCTGGACT CATGTCTACC TTGCCTACCG TTGATGGAAA AGGTTTTGAC ATGCTAGACC 2400

TTGGTGCCAA TGCAGAAAAT ACAGCCCAGC ACCTCCATCA ATATGCGGTT CTAGGTTCCT 2460

TCTATGCTAA AAATGTCCGT GGCATTGCGC AACCACGCGT TGGTTTGCTC AACAACGGAA 2520

CAGAGAGTAG CAAGGGCGAC CCGCTTCGTA AGGAAACTTA TGAATTACTG GCGGCTGATG 2580

AAAGTTTGAA CTTTATCGGA AACGTGGAAG CGCGTGATTT GATGAATGGC GTTGCAGATG 2640

TTGTTGTGGC AGATGGTTTC ACGGGAAACG CTGTGCTCAA ATCCATCGAA GGGACAGCTA 2700

TGGGAATCAT GGGCTTGCTC AAGACAGCTA TTACAGGTGG TGGTCTTCGA GCGAAACTAG 2760

GTGCCCTCCT TCTCAAGGAC AGCCTCAGTG GTTTGAAAAA ACAGCTCAAT TATTCAGATG 2820

TTGGTGGAGC GGTCTTGTTT GGTGTTAAGG CACCTGTTGT CAAGACTCAT GGCTCAAGCG 2880

ATGCCAAGGC TGTTTATAGT ACGATTCGTC AGATCCGTAC CATGCTAGAA ACAGACGTGG 2940

TTGCCCAGAC TGCGCGTGAA TTTTCAGGAG AATAAAAGAG ATGACAGAAA AAGAAATTTT 3000

TGACCGTATT GTGACCATTA TCCAAGAGCG ACAGGGAGAG GACTTTGTCG TGACAGAATC 3060

CTTGAGTCTG AAAGACGATT TGGATGCGGA TTCTGTTGAC TTGATGGAGT TTATCTTGAC 3120

TCTGGAAGAT GAATTTAGTA TCGAAATCAG CGATGAAGAA ATTGACCAAC TCCAAAACG? 3180

AGGAGATGTG GTTAAAATCA TTCAAGGAAA ATAGCAATCG GAGTTCCAAG TCAACGGAAG 3240

TAGATGGTTT TTAGAAATGA GAAATATCGG ACAAGCTGGT AAAATCTTGG CTGACAGTGG 3300

TTATCAAGGG CTCATGAAGA TATATCCTCA AGCACAAACT CCACGTAAAT CCAGCAAACT 3360

CAAGCCGCTA ACAGTTGAAG ATAAAGCCTG TAATCATGCG CTATCTAAGG AGATAAGCAA 3420

GGTTGAGAAT ATCTTTGCCA AAGTAAAAAC GTTTAAAATG TTTTCAACAA CCTATCGAAA 3480

TCATCGTAAA CGCTTCGGAT TACGAATGAA TTTGATTGCT GGTATTATCA ATCATGAACT 3540

AGGATTCTAG TTTTGCAGGA AGTCTAATAG TAAAAAAGTG ATTAGAAAAC ATCTTTTTTA 3600

AAAATAGAGA TGATTTTGAA ACAAAAAAGC TAATTCAAGA CGTTTCGATG CCAATTCAAG 3660

ATTTGGATGA AAAAAATTAA TAGATACTGT TATACTAAAC TTGTCAAGTT TGTAACAAGA 3720

CAAATATTAA AAATAAAAAA GAGGTATTCG TTATGAATAC AAAAACGATG TCACAATTTG 3780

AAATTATGGA TACTGAGATG CTTGCTTGCG TTGAAGGTGG CGGATGCAAT TGGGGAGATT 3840

TTGCCAAAGC AGGTGTTGGA GGAGGAGCAG CACGAGGTCT TCAGCTAGGA ATTAAAACAA 3900

GAACATGGCA AGGTGCAGCA ACTGGTGCTG TGGGAGGAGC TATACTTGGA GGTGTGGCCT 3960

ATGCAGCGAC ATGTTGGTGG TAATTATGGA TTTTAAAAGT TTTATTATTG GTTTAGTAGT 4020

TGGTATATTT GGTCCTTATA TGGATGATTT AATTAGAAAA AAATTTTTAA AGTCTTCGGA 4080

GAAGAAAACA GAAAAATCTG TTAAAAAATA ATCAAAACTA TAAATGATGA ATCTGAATCA 4140

AAATTATTTT GCGCATGTAA AGAGGAGTCT TATAGTAACG AGTCAAAAAA GGAGTAACTA 4200

TGAATCGTAA TTTAGAACGG TGTTATCTAT TCTGACTAGG AATAGATCAT ACCAGAGGTA 4260

GCTTAGAAAT AGCAGAGACA TTAGAAATTG AAGTAATAAA TAGGATGTCG TAAGTGTTAC 4320

TATCAATGAT TTATTTGTTT CAAGCTTGCC TAGGGTGACA GTAAAAAATC AATTTCCTTT 4380

CAATAGCATA TTTTTAGTGG GCAGGACTCT TGTTCTGCCT ATTTTTTTAT CCAAAAAGTG 4440

CAGTTGGGAG GGAGATAGGC TCATTTGGGA AGGAAGTCCA GTTTTTGTTT AGTGATTGGG 4500

GTAAGATAGT TGTTATCAGA TGAGTTAATA CTCTTCGAAA ATCAAATTCA AACCACGTCA 4560

ACGTCGCCTT GCCGTATATA TGTGACTGAC TTCGTCAGTC CTATCTACAA CCTCAAAACA 4620

GTGTTTTGAG CAGCCTACGG CTAGTTTCCT AGTTTGCTCT TTGATTTTCA TTGAGTATTA 4680

GGGAAAAGGA GATGAATATG AAATTTGGGA AACGTCATTA TCGTCCGCAG GTGGATCAGA 4740

TGGACTGCGG TGTAGCTTCA TTAGCCATGG TTTTTGGCTA CTATGGTAGT TATTATTTTT 4800

TGGCTCACTT GCGAGAATTG GCTAAGACGA CCATGGATGG GACGACGGCT TTGGGCTTGG 4860

TCAAGGTGGC AGAGGAGATT GGTTTTGAGA CGCGAGCCAT TAAGGCAGAT ATGACGCTTT 4920

TTGACTTGCC GGATTTAACT TTTCCTTTTG TTGCCCATGT GCTTAAGGAA GGGAAATTGC 4980

TCCACTACTA TGTGGTGACT GGGCAGGATA AGGATAGCAT TCATATTGCC GATCCAGATC 5040

CCGGGGTGAA GTTGACTAAA CTGCCACGTG AGCGTTTTGA GGAAGAATGG ACAGGAGTGA 5100

CTCTTTTTAT GGCACCTAGT CCAGACTATA AGCCTCATAA GGAACAAAAA AATGGTCTGC 5160

TCTCTTTTAT CCCTATATTA GTGAAGCAGC GTGGCTTGAT TGCCAATATC GTTTTGGCAA 5220

CACTCTTGGT AACCGTGATT AACATTGTGG GTTCTTATTA TCTGCAGTCT ATCATTGATA 5280

CCTATGTGCC AGATCAGATG CGTTCGACAC TAGGGATTAT TTCTATTGGG CTAGTCATCG 5340

TCTACATCTT CCAGCAAATC TTGTCTTACG CTCAGGAGTA TCTCTTGCTT GTTTTGGGGC 5400

AACGCTTGTC GATTGACGTG ATTTTGTCCT ATATCAAGCA TGTTTTTCAC CTCCCTATGT 5460

CCTTCTTTGC GACACGCAGG ACAGGGGAGA TCGTGTCTCG TTTTACAGAT GCTAACAGTA 5520

TCATCGATGC GCTGGCTTCG ACCATCCTTT CGATTTTCCT AGATGTGTCA ACGGTTGTCA 5580

TTATTTCCCT TGTTCTATTT TCACAAAATA CCAATCTCTT TTTCATGACT TTATTGGCGC 5640

TTCCTATCTA CACAGTGATT ATCTTTGCCT TTATGAAGCC GTTTGAAAAG ATGAATCGGG 5700

ATACCATGGA AGCCAATGCG GTTCTGTCTT CTTCTATCAT TGAGGACATC AACGGTATTG 5760

AGACTATCAA GTCCTTGACC AGTGAAAGTC AGCGTTACCA AAAAATTGAC AAGGAATTTG 5820

TGGATTATCT GAAGAAATCC TTTACCTATA GTCGAGCAGA GAGTCAGCAA AAGGCTCTGA 5880

AAAAGGTTGC CCATCTCTTG CTTAATGTCG GCATTCTCTG GATGGGGGCT GTTCTGGTCA 5940

TGGATGGCAA GATGAGTTTG GGGCAGTTGA TTACCTATAA TACCTTGCTG GTTTACTTTA 6000

CTAATCCTTT GGAAAATATC ATCAATCTGC AAACCAAGCT TCAGACAGCG CAGGTTGCCA 6060

ATAACCGTCT AAATGAAGTG TATCTAGTAG CTTCTGAGTT TGAGGAGAAG AAAACAGTTG 6120

AGGATTTGAG CTTGATGAAG GGAGATATGA CCTTCAAGCA GGTTCATTAC AAGTATGGCT 6180

ATGGTCGAGA TGTCTTATCG GATATCAATT TAACCGTTCC CCAAGGGTCT AAGGTGGCTT 6240

TTGTGGGGAT TTCAGGGTCA GGTAAGACGA CTTTGGCCAA GATGATGGTT AATTTTTACG 6300

ACCCAAGTCA AGGGGAGATT AGTCTGGGTA GTGTCAATCT CAATCAGATT GATAAAAAAG 6360

CCCTGCGCCA GTACATCAAC TATCTGTCTC AACAGCCCTA TGTCTTTAAC GGAACGATTT 6420

TGGAGAATCT TCTTTTGGGA GCCAAGGAGG GGACGACACA GGAAGATATC TTACGGGCGG 6480

TCGAATTGGC AGAGATTCGA GAGGATATCG AGCGCATGCC ACTGAATTAC CAGACAGAAT 6540

TGACTTCGGA TGGGGCAGGG ATTTCAGGTG GTCAACGTCA GAGAATCGCT TTGGCGCGTG 6600

CTCTCTTGAC AGATGCGCCG GTCTTGATTT TGGATGAGGC GACTAGCAGT TTGGATATTT 6660

TGACAGAGAA GCGGATTGTC GATAATCTCA TTGCTTTGGA CAAGACCTTG ATTTTCATTG 6720

CTCACCGCTT GACTATTGCT GAGCGGACAG AGAAGGTAGT TGTCTTGGAT CAGGGCAAGA 6780

TTGTCGAAGA AGGAAAGCAT GCTGATTTGC TTGCACAGGG TGGCTTTTAC GCCCATTTGG 6840

TCAATAGCTA GAAAGAGGAG AGGATGAAAC CAGAATTTTT AGAAAGTGCG GAGTTTTATA 6900

ATCGTCGTTA CCATAATTTT TCCAGTAGTG TGATTGTACC CATGGCCCTT CTGCTTGTGT 6960

TTTTACTTGG CTTTGCAACT GTTGCAGAGA AGGAGATGAG TTTGTCCACT AGAGCTACTG 7020

TCGAACCTAG TCGTATCCTT GCAAATATCC AGTCAACTAG CAACAATCGT ATTCTTGTCA 7080

ATCATTTGGA AGAAAATAAG CTGGTTAAGA AGGGGGATCT TTTGGTTCAA TACCAAGAAG 7140

GGGCAGAGGG TGTCCAAGCG GAGTCCTATG CCAGTCAGTT GGACATGCTA AAGGATCAAA 7200

AAAAGCAATT GGAGTATCTG CAAAAGAGCC TGCAAGAAGG GGAGAACCAC TTTCCAGAGG 7260

AGGATAAGTT TGGCTACCAA GCCACCTTTC GCGACTACAT CAGTCAAGCA GGCAGTCTTA 7320

GGGCTAGTAC ATCGCAACAA AATGAGACCA TCGCGTCCCA GAATGCAGCA GCTAGCCAAA 7380

CCCAAGCCGA AATCGGCAAC CTCATCAGTC AAACAGAGGC TAAAATTCGC GATTACCAGA 7440

CAGCTAAGTC AGCTATTGAA ACAGGTGCTT CCTTGGCCGG TCAGAATCTA GCCTACTCTC 7500

TTTACCAGTC CTACAAGTCT CAGGGCGAGG AAAATCCCCA AACTAAGGTT CAGGCAGTTG 7560

CACAGGTTGA AGCACAGATT TCTCAGTTAG AATCTAGTCT TGCTACTTAC CGTGTCCAGT 7620

ATGCAGGTTC AGGTACCCAG CAAGCCTATG CGTCAGGGTT AAGCAGTCAA TTGGAATCCC 7680

TTAAATCCCA ACACTTGGCA AAGGTTGGTC AGGAATTGAC CCTTCTAGCC CAGAAAATTT 7740

TGGAGGCAGA GTCAGGTAAG AAGGTACAGG GAAATCTTTT AGACAAGGGG AAAGTTACGG 7800

CGAGTGAGGA TGGGGTGCTT CATCTTAATC CTGAGACCAG TGATTCTAGC ATGGTTGCAG 7860

AAGGTGCCCT ACTAGCCCAA CTTTATCCAT CTTTGGAAAG AGAAGGGAAA GCCAAACTCA 7920

CAGCTTATCT AAGTTCAAAA TATGTAGCAA GAATCAAGGT CGGTGATTCT GTTCGCTATA 7980

CTACGACTCA TGATGCCGGG AATCAACTTT TCCTAGATTC TACTATTACA AGTATTGATG 8040

CGACAGCTAC TAAGACTGAG AAAGGGAATT TCTTTAAAAT CGAGGCGGAG ACTAATCTAA 8100

CTTCGGAGCA GGCTGAAAAA CTTAGGTACG GGGTGGAAGG CCGCTTGCAG ATGATTACGG 8160

GCAAGAAAAG TTACCTACGT TATTATTTGG ATCAATTTTT GAACAAAGAG TAATGTTCGT 8220

GTTTTTAGAG TTAAATAATT TTTAAACTGT GAGAAAGATT CTTCTTGCAG TTTTTTCTTT 8280

ACAATTTTTG AAAAACATCT ACTATTTATT CGGTTAAATT CTTGTGTTTT TTGGTTTTTT 8340

GTGGTAAAAT GTGCTCAAGT AATACGAAAG GCGAACTTTA AAATGTCAAA ACAATTGATC 8400

TATTCGGGAA AAGCTAAAGA TATCTATACA ACTGAGGATG AAAATCTTAT TATTTCAACT 8460

TACAAGGACC AGGCGACTGC TTTCAACGGT GTCAAGAAGG AGCAGATTGC AGGTAAGGGA 8520

GTCTTGAATA ATCAGATCTC ATCTTTTATT TTTGAGAAAT TAAATGTGGC TGGTGTGGCG 8580

ACTCACTTTG TGGAGAAACT TTCAGACACG GAACAACTCA ATAAAAAGGT TAAGATTATT 8640

CCTTTGGAAG TCGTGCTCCG CAACTATACT GCTGGTTCCT TTTCAAAACG TTTTGGTGTG 8700

GATGAGGGAA TCGCCTTGGA GACTCCGATT GTCGAATTTT ACTACAAAAA TGATGATTTG 8760

GATGATCCAT TTATCAATGA TGAGCATGTG AAATTCCTAC AGATTGCGGG TGACCAGCAG 8820

ATTGCCTACT TGAAGGAAGA AACGCGTCGT ATCAATGAAC TATTGAAAGT CTGGTTTGCT 8880

GAGATTGGGC TTAAATTGAT TGACTTTAAG CTAGAGTTCG GTTTTGACAA GGATGGCAAG 8940

ATTATCTTGG CAGACGAATT TTCACCAGAT AACTGCCGCT TGTGGGACGC TGATGGCAAC 9000

CACATGGATA AGGATGTTTT CCGTAGAGGA TTGGGAGAAC TAACCGACGT TTATGAGATT 9060

GTTTGGGAAA AGTTGCAGGA ATTGAAATAA TCTGTTTGCA ACGGAAAACC TTCGTCTCTC 9120

AACTAAAAGG ACTCAGGCTG AAAAGGTCCC CCAGACCTTT TCACTCTGTA GAGAACTAGG 9180

TGAACTAACA GATGTTTACG AAATTGTCTG GGAAAAGTTG CAGGGTTTAA AATAACAACC 9240

TCAAGGCTGT TTGGGAATAT TGCAAGAGCT GAAATAAAGG AATAAGAATT GATGGATAAA 9300

CGTATTTTTG TTGAAAAAAA GGCTGATTTT CAGGTCAAGT CAGAGAGTTT GGTTAGAGAG 9360

CTCCAGCACA ACTTGGGACT GTCAAGCTTG AAAAGTATTC GTATTGTGCA AGTATATGAT 9420

GTATTTGACT TGGCTGAGGA CTTGTTTGCA CCTGCAGAGA AGCACATTTT CTCTGAGCAG 9480

GTAACCGACC ATGTTTTAGA TGAAGTATCT GTGCAGGCGG ATCTTGCTAA CTATGCTTTC 9540

TTTGCCATTG AAAGTCTGCC AGGGCAGTTT GACCAGCGTG CAGCTTCGTC ACAGGAAGCC 9600

TTGCTTTTGT TGGGAAGTTC GAGTGACGTG ACAGTCAACA CAGCCCAACT TTACTTGGTG 9660

AATAAAGATA TTGATGCGAC TGAGTTGGAA GCTGTCAAAA ACTACCTGCT CAATCCAGTT 9720

GATTCTCGTT TCAAGGATAT CACGACAGGG ATTGCCAAGC AGGAGTTTTC AGAGTCAGAC 9780

AAGACCATTC CCAAATTGAC TTTCTTTGAA AGCTATGCAG CAGAAGACTT TGCTCGCTAC 9840

AAGGCCGAAC AAGGGATGGC CATGGAAGTG GATGATTTGC TCTTTATCCA AGACTACTTT 9900

AAGTCAATCG GGCGCGTGCC AACTGAGACT GAACTCAAGG TTTTGGACAC TTACTGGTCT 9960

GACCACTGCC GTCATACGAC TTTTGAGACA GAGTTGAAAC ACATCGACTT TTCAGCTTCT 10020

AAATTTCAAA AGCAATTGCA GTCAACCTAT GACAAGTATA TTGCCATGCG CGAGGAATTA 10080

GGTCGGTCTG AAAAACCACA AACCTTGATG GATATGGCGA CTATTTTCGG TCGTTATGAG 10140

CGTGCTAATG GACGATTGGA TGATATGGAA GTCTCTGACG AAATCAATGC CTGCTCAGTT 10200

GAAATTGAAG TGGACGTTGA TGGTGTCAAG GAACCTTGGC TCCTCATGTT TAAAAACGAA 10260

ACCCACAACC ATCCAACAGA AATTGAGCCA TTTGGTGGAG CGGCTACCTG TATTGGTGGA 10320

GCTATTCGTG ATCCGTTGTC AGGCCGTTCC TATGTTTACC AAGCCATGCG TATTTCAGGT 10380

GCTGGTGATA TTACAGCACC GATTTCGGAA ACTCGCGCTG GGAAATTGCC ACAACAAGTC 10440

ATTTCTAAAA CAGCAGCTCA TGGTTATTCT TCATATGGTA ACCAGATTGG GCTTGCAACA 10500

ACCTACGTTC GTGAATACTT CCACCCAGGC TTTGTAGCTA AACGTATGGA ACTTGGTGCC 10560

GTTGTTGGTG CGACTCCCAA GGGCAATGTT GTCCGTGAAA AACCTGAAGC AGGTGATGTG 10620

ATCATCCTTC TCGGAGGCAA AACAGGTCGT GATGGTGTCG GTGGTGCGAC GGGCTCTTCT 10680

AAGGTTCAAA CAGTTGAGTC TGTAGAGACT GCTGGTGCTG AGGTTCAAAA AGGAAATGCC 10740

ATCGAAGAAC GCAAGATTCA GCGCCTCTTC CGTAATGGCA ATGTCACTCG TCTGATCAAG 10800

AAGTCCAATG ACTTTGGGGC AGGCGGTGTC TGTGTGGCTA TCGGTGAATT GGCAGACGGT 10860

CTTGAAATCG ACCTCAACAA GGTGCCTCTT AAATACCAGG GCTTGAATGG TACAGAAATT 10920

GCCATCTCTG AATCACAAGA ACGGATGGCG GTCGTGGTTC GTCCTGAAGA TGTGGATGCC 10980

TTCGTTGCCG AATGTAACAA AGAAAATATT GATGCTGTTG TGGTGGCGAC AGTAACTGAA 11040

AAACCAAATC TTGTCATGCA CTGGAATGGT GAGACAATCG TTGACTTGGA GCGTCGTTTC 11100

CTTGACACCA ATGGTGTGCG CGTGGTTGTC GATGCCAAAG TTGTGGACAA GGATGTCAAA 11160

CTCCCAGAAG AGCGTCAAAC ATCTGCTGAA ACACTGGAAT CAGATACCCT TACGGTTCTA 11220

TCTGACCTCA ACCATGCAAG TCAAAAAGGA TTACAGACTA TCTTTGACTG CTCTGTTGGA 11280

CGCTCAACGG TTAATCACCC ACTTGGTGGT CGTTACCAAC TCACACCAAC TGAGGCATCT 11340

GTGCAGAAAT TGCCAGTTCA ACACGGTGTG ACTCATACTG CGTCGGTCAT TGCTCAAGGT 11400

TTCAACCCAT ATGTAGCTGA ATGGTCTCCA TACCACGGTG CTGCTTATGC GGTTATCGAA 11460

GCAACTGCTC GTTTGGTGGC TGCTGGTGCC AACTGGTTCA AGGCTCGTTT CTCTTACCAA 11520

GAGTATTTCG AGCGTATGGA TAAACAAGCA GAGCGTTTCG GTCAGCCAGT AGCTGCTCTT 11580

CTAGGTTCTA TTGAAGCACA AATTCAGCTT GGCTTGCCAT CTATCGGTGG TAAGGACTCC 11640

ATGTCTGGTA CCTTTGAAGA ATTGACCGTT CCGCCAACCT TGGCTGCCTT TGGGGTGACG 11700

ACGGCAGATA GCCGTAAGGT GCTCTCTCCA GAATTTAAAG CTGTTGGGGA AAATATCTAC 11760

TACATCCCAG GTCAAGCCCT CTCTGCAGAG ATTGATTTTG ACTTGATTAA GAAAAATTTT 11820

GCTCAGTTTG AAGCCATCCA AGCTGACCAT AAAGTGACAT CTGCATCAGC TGTCAAATAC 11880

GGTGGTGTAG TTGAAAGTTT GGCTCTTGCT ACCTTTGGAA ACTATATTGG TGCAGAGGTG 11940

ACCTTGCCTG AACTTGAAAC AGCTTTGACA GCTCAATTAG GCGGCTTTGT CTTCACATCT 12000

CCTGAAGAAA TTGCTGGAGT AGAGAAGGTT GGACAAACGA AAGCAGACTT TACACTGACT 12060

GTCAACGGTG TGAAGCTAGA TGGACACAAG CTTGACAGTG CATTTCAAGG GACATTGGAA 12120

GAAGTTTACC CAACAGAATT TACCCAAGCG AAAGAACTAG AAGAAGTACC AGCTGTGGCA 12180

TCAGATGTTG TGATTAAAGC CAAAGAAAAG GTTGAAAAAC CTGTGGTTTA CATCCCAGTC 12240

TTTCCAGGAA CCAACTCAGA ATATGATTCA GCTAAGGCCT TCGAAAAAGA AGGTGCAGAG 12300

GTCAATTTGG TGCCATTCGT GACCTTGAAT GAAGAAGCTA TTGTCAAGTC AGTTGAAACT 12360

ATGGTTGACA ATATCGACAA GACTAATATT CTCTTCTTTG CTGGTGGATT CTCGGCTGCG 12420

GATGAACCAG ATGGTTCAGC TAAGTTTATC GTCAATATCC TGCTTAATGA AAAAGTGCGT 12480

GTGGCTATTG ATAGCTTTAT CGCCCGTGGT GGTTTGATTA TCGGTATTTG TAATGGATTC 12540

CAAGCCTTAG TCAAATCGGG TCTCCTACCC TACGGAAACT TTGAAGCTGC TAACAGTACT 12600

AGCCCAACCC TCTTCTACAA TGATGCCAAC CAACACGTGG CCAAGATGGT GGAAACTCGC 12660

ATTGCCAATA CCAACTCACC ATGGTTGGTT GGTGTGCAAG TGGGCGATAT CCACGCTATT 12720

CCTGTTTCGC ACGGTGAAGG GAAGTTTGTC GTGACGGCTG AGGAATTTGC AGAGCTCCGT 12780

GACAATGGAC AAATTTTCAG CCAATACGTT GACTTTAACG GTAAACCAAG TATGGATTCT 12840

AAGTACAATC CGAATGGTTC TGTCCATGCC ATCGAAGGAA TTACCAGCAA GAATGGTCAA 12900

ATCATCGGTA AGATGGGCCA CTCAGAACGT TATGAGGATG GTCTTTTCCA AAATATCCCA 12960

GGCAATAAAG ACCAACACCT GTTCGCATCA GCGGTTAAAC ATTTCACTGG AAAATAAGAC 13020

TTACAGATTT TCTAATAGAT AGTATCAGTA ATGTAAAAGT CATGTAAATC TAGCTCTTGA 13080

TGATTACAAA TGAAAATTAG GTATAAAAAA TGACATACGA AGTAAAATCT CTTAATGAAG 13140

AATGTGGTGT TTTCGGTATT TGGGGACATC CAGATGCTGC TAAGTTGACC TATTTTGGAC 13200

TCCACAGTCT TCAACACCGT GGTCAGGAGG GGGCAGGAAT CCTCTCCAAT GATCAAGGAC 13260

AACTGAAGCG CCATCGTGAC ATGGGGCTTT TATCAGAAGT TTTCAGAAAT CCAGCTAATT 13320

TGGATAAATT GACAGGAGCT GGTGCGATTG GGCATGTGCG TTATGCGACT GCTGGCGAAG 13380

CTTCTGTAGA TAACATCCAG CCCTTCCTCT TCCGTTTTCA CGATATGCAG TTTGGTTTGG 13440

CTCATAATGG AAATCTGACC AATGCAGCCT CTCTCAAGAA AGAACTGGAA CAAAGAGGAG 13500

CAATTTTCAG CGCGACTTCG GACTCTGAAA TCTTGGCTCA CCTCATTCGT CGCAGTCATA 13560

ATCCTAGCCT GATGGGCAAA ATCAAGGAAG CGCTCAGCCT TGTCAAAGGT GGTTTTGCCT 13620

ATATCTTGCT GTTTGAGGAC AAGTTGATTG CGGCTCTTGA CCCAAATGGA TTCCGACCGC 13680

TTTCGATTGG TAAAATGGCT AATGGAGCAG TTGTTGTATC TTCTGAAACC TGTGCTTTTG 13740

AGGTCATTGG TGCCGAGTGG ATTCGTGATT TGAAGCCAGG TGAGATTGTG ATCATTGATG 13800

ACGAGGGCAT TCAGTATGAC AGCTATACAG ATGATACCCA GTTGGCGGTT TGTTCTATGG 13860

AGTATATCTA CTTTGCTCGC CCTGATTCTA ATATCCACGG TGTCAATGTC CATACGGCAC 13920

GTAAGAGAAT GGGAGCGCAA TTGGCGCGAG AATTTAAGCA TGAGGCAGAT ATTGTAGTTG 13980

GTGTGCCCAA TTCTTCCCTA AGCGCGGCTA TGGGATTTGC GGAAGAATCA GGCTTACCAA 14040

ATGAAATGGG TCTGATCAAA AACCAATACA CCCAGCGAAC TTTTATCCAA CCGACTCAAG 14100

AATTGCGGGA GCAAGGAGTG CGGATGAAAC TGTCTGCTGT TTCGGGTGTT GTCAAAGGCA 14160

AACGTGTGGT CATGGTGGAT GATTCCATTG TACGTGGAAC AACCTCTCGT CGTATCGTTC 14220

AGCTCTTGAA AGAAGCGGGT GCGACTGAGG TTCACGTTGC CATTGGAAGT CCTGCACTAG 14280

CGTATCCATG TTTCTACGGG ATTGATATCC AGACCCGTCA GGAGCTGATT GCAGCCAATC 14340

ATACGGTCGA AGAAACTCGC CAAATCATTG GTGCGGACAG TCTGACTTAT CTTTCAATTG 14400

ATGGCTTGAT TGAGTCGATT GGTATCGAAA CAGATGCGCC GAACGGTGGT CTCTGTGTCG 14460

CTTACTTTGA CGGTGACTAC CCAACGCCTC TTTATGACTA CGAAGAAGAC TATCGTAGAA 14520

GTTTGGAAGA AAAGACCAGT TTTTACAAGT AGGCGACAGA TTCTCCATTA AAGAAAAGGA 14580

AAAAATAAAT GACAAATAAA AATGCATATG CCTCACGTCT CACTACTGAC TAAAGGCTTA 14640

AGCATTTAGT CAGTAGACGC TTTGTCCTAT AGGATCAAAG CTAGAGCCCT GACTAGTATT 14700

TTTAGATAAA AAGATGGTTT ATCTAAAAAT ACGTCGCAGT CTTTCTCAAA AAAAGAAAAG 14760

GAAAAATAAA ATGGCAAATA AAAATGCGTA CGCTCAATCT GGTGTGGATG TTGAAGCGGG 14820

TTATGAAGTT GTTGAACGGA TTAAAAAGCA CGTGGCCCGT ACGGAGCGTG CAGGTGTCAT 14880

GGGAGCTCTT GGTGGCTTTG GTGGTATGTT TGACCTTTCC AAGACTGGGG TTAAAGAACC 14940

CGTCTTGATT TCAGGGACTG ACGGTGTCGG AACCAAGCTC ATGTTGGCTA TCAAGTACGA 15000

CAAGCACGAT ACCATCGGGC AGGACTGTGT GGCCATGTGT GTCAACGACA TCATTGCTGC 15060

AGGTGCGGAA CCCCTCTATT TTCTCGACTA CGTAGCGACA GGGAAGAATG AACCAGCTAA 15120

GCTAGAACAA GTGGTTGCTG GTGTGGCAGA AGGTTGTGTG CAGGCTGGTG CTGCCCTCAT 15180

CGGTGGGGAA ACGGCTGAAA TGCCGGGCAT GTACGGCGAA GACGACTATG ACTTGGCTGG 15240

TTTTGCGGTC GGTGTGGCTG AAAAATCTCA AATCATTGAC GGTTCAAAGG TGGTAGAGGG 15300

AGATGTTCTT CTCGGACTTG CTTCAAGTGG GATTCACTCA AATGGTTACT CTTTGGTTCG 15360

TCGTGTCTTT GCGGATTACA CAGGTGAGGA AGTCCTACCA GAATTGGAAG GCAAGAAACT 15420

TAAGGAAGTT CTACTTGAGC CGACTCGTAT CTATGTCAAG GCTGTCTTGC CGCTCATCAA 15480

AGAAGAGTTG GTCAACGGCA TTGCCCACAT CACAGGTGGT GGCTTTATCG AAAATGTCCC 15540

TCGTATGTTT GCAGATGACC TAGCTGCTGA AATTGATGAA AGTAAAGTTC CAGTGCTTCC 15600

AATTTTCAAA ACCCTTGAAA AATACGGTCA GATTAAACAC GAAGAAATGT TTGAAATCTT 15660

CAATATGGGT GTGGGACTTA TGTTGGCGGT CAGCCCTGAA AATGTAGAGC GTGTAAAAGA 15720

ATTGTTGGAT GAAGCAGTCT ATGAAATTGG TCGCATCGTC AAGAAAGAAA ACGAAAGTGT 15780

CATTATCAAA TGAAAAAAAT AGCGGTTTTT GCCTCTGGTA ATGGCTCAAA TTTTCAGGTG 15840

ATTGCCGAAG AATTTCCAGT GGAGTTTGTC TTTTCAGACC ATCGTGATGC CTATGTGCTT 15900

GAGCGTGCAA AGCAGCTCGG CGTTCTGTCC TATGCTTTTG AACTCAAGGA GTTTGAGAGC 15960

AAGGCAGACT ACGAAGCAGC CCTTGTCGAA CTCTTGGAAG AACACCAGAT TGACTTGGTT 16020

TGCCTAGCAG GCTACATGAA AATCGTTGGA CCAACCTTAT TGTCGGCTTA TGAAGGTCGG 16080

ATTGTCAACA TTCATCCAGC CTACTTGCCA GAATTTCCAG GAGCTCATGG GATTGAGGAT 16140

GCTTGGAATG CTGGCGTGGG TCAGTCTGGT GTGACCATTC ACTGGGTGGA TTCGGGTGTG 16200

GATACAGGCC AGGTCATCAA ACAGGTTCGT GTGCCACGAC TAGCTGATGA TACCATTGAC 16260

AGATTTGAAG CTCGCATCCA TGAAGCAGAG TACAGGCTGT ATCCGGAAGT AGTGAAGGCT 16320

CTATTTACAG ATTGACTTTT TGATGATTCA TATGATATCT TTGATTTTAA ATTGGAGTCA 16380

GTGTTTGTTG AAGACGGCTT CAAACGGAGG TATTTGTAAT GTTAGAATCT AAAAAAACAA 16440

CTCGATATGT ATTTTATGTC TATCTGATGT TATTAACTTG GGGAATCTTA TTTAAGTTTG 16500

AAACAAATCC TGAATTTATA GCATTTTTCT TAGCTCCAAG GTATATCAAT TGGATTCCAT 16560

TTTCAGAACC ACTAATAGTC GATGGAAAAA TTGTTTTTGC TGAAATGTTA TTTAATCTGA 16620

TTTTCTTTAT TCCATTAGGT GTTTGTTTCC CTTTGATAAA AACTAATTTA TCTAGTTTAA 16680

GAATAGTCGG GACAGGTTTC TTGATTAGTT TATTGTTTGA GTGCTTACAG TATATTTTAG 16740

CAATAGGTAT AACAGATATA ACGGATTTGA CTTTAAATAC GCTAGGTGTC TGTGTAGGCT 16800

TACTGATTTA TCAAATTTTT ATAAGAGTGT TCAAATCACA GACTAGAAAA TGGATCAATA 16860

TCTTAGGTAT GCTTAGCCTT GGTTTTGCTT ATCTTGTTTT ACTGTTACTG CATTTACTTA 16920

GTGTTTAACT AATGATTAAA AAGGAGAATA TAATGACTAA ACGCGTCTTA ATCAGCGTCT 16980

CAGACAAAGC GGGCATTGTT GAATTTGCCC AAGAACTCAA AAAACTTGGT TGGGAGATTA 17040

TCTCAACAGG TGGAACTAAG GTTGCCCTTG ATAATGCTGG GGTGGATACC ATTGCTATCG 17100

ATGATGTGAC TGGTTTCCCA GAAATGATGG ACGGTCGTGT GAAGACCCTC CACCCAAATA 17160

TCCACGGAGG GCTTCTCGCT CGTCGTGACT TGGATAGCCA CTTGGAAGCG GCTAAGGACA 17220

ACAAGATTGA GCTCATTGAC CTTGTGGTGG TCAACCTTTA CCCATTTAAG GAAACTATCC 17280

TTAAACCAGA TGTGACTTAT GCTGATGCAG TTGAAAATAT CGATATTGGT GGGCCATCTA 17340

TGCTTCGTTC AGCAGCGAAA AATCATGCCA GTGTTACAGT TGTGGTAGAT CCTGCTGACT 17400

ACGCTGTGGT TTTGGATGAA TTGGCAGCAA ACGGCGAAAC CTCTTATGAA ACTCGCCAAC 17460

GTTTAGCAGC CAAAGTATTT CGTCACACAG CGGCTTATGA CGCCTTGATT GCAGAATACT 17520

TCACAGCTCA AGTGGGTGAA AGCAAGCCTG AAAAACTCAC TTTGACTTAT GACCTCAAGC 17580

AACCAATGCG TTACGGTGAG AATCCTCAAC AAGACGCGGA CTTTTACCAG AAAGCTTTGC 17640

CTACAGACTA CTCCATTGCT TCAGCCAAAC AGCTCAACGG GAAAGAATTG TCATTTAATA 17700

ATATCCGTGA TGCAGATGCT GCTATCCGTA TCATCCGTGA CTTCAAAGAT AGTCCAACCG 17760

TTGTGGCTCT CAAACACATG AATCCATGTG GAATTGGTCA AGCTGATGAC ATCGAGACTG 17820

CTTGGGACTA CGCTTATGAG TCTGACCCAG TGTCTATCTT TGGTGGGATT GTCGTCCTCA 17880

ACCGTGAGGT GGATGCTGCG ACAGCTGAGA AGATGCACGG CGTTTTCCTC GAAATCATCA 17940

TTGCACCAAG CTATACGGAT GAAGCGCTAG CCATTTTGAT CAATAAAAAG AAAAACTTGC 18000

GTATCCTTGC CTTGCCATTT AATGCTCAAG AGGCTAGCGA AGTGGAAGCA GAATACACAG 18060

GTGTAGTCGG TGGACTTCTC GTGCAAAATC AAGACGTGGT CAAGGAAAGC CCAGCTGACT 18120

GGCAAGTGGT GACTAAACGT CAGCCAACTG AGACAGAAGC GACTGCTCTT GAGTTCGCTT 18180

GGAAGGCTAT CAAGTACGTC AAATCAAATG GTATTATCGT GACCAACGAC CACATGACAC 18240

TTGGTGTTGG TCCAGGTCAA ACCAACCGTG TGGCTTCTGT TCGCCTTGCC ATTGACCAAG 18300

CCAAAGATCG TCTGGACGGG GCGGTCCTTG CTTCAGATGC CTTCTTCCCA TTTGCGGATA 18360

ACGTGGAAGA AATCGCCAAA GCAGGAATTA AGGCCATCAT CCAGCCCGGT GGCTCTGTCC 18420

GTGACCAAGA ATCCATCGAA GCAGCGGATA AATACGGCTT GACTATGGTC TTTACAGGTG 18480

TGAGACATTT TAGACATTAA GAAGATAAAA GGGAAGAAAA CAGTTTCTTT CCTTTTTTGG 18540

CTTAAAATAC TAACTGAAAC AAGATTAAAA CGAACTTTTT TGATATAATG TTGGTAAATA 18600

ATTCGCAAAA GAGGTTGAGG AATGAAACTG CTTGTTGTCG GTTCTGGTGG TCGTGAGCAT 18660

GCGATTGCTA AAAAGTTACT TGAATCAAAA GACGTGGAAA AAGTCTTTGT AGCTCCTGGG 18720

AATGATGGGA TGACTCTGGA TGGTTTGGAA TTGGTAAATA TCTCTATTTC CGAACATTAT 18780

AAATTGATTG ACTTCGCAAA GACCAATGAT GTTGCTTGGA CCTTTATCGG TCCAGATGAT 18840

GCCCTTGCTG CTGGTATCGT GGATGATTTT AACCAAGCTG GACTTAAGGC CTTTGGTCCG 18900

ACTAGGGCTG CAGCGGAGCT GGAGTGGTCC AAGGATTTCG CCAAGGAAAT CATGGTCAAA 18960

TACGGCGTTC CGACAGCAAC ATATGGCACA TTTTCAGATT TCGAGGAAGC CAAAGCCTAT 19020

ATCGAAAAGC ATGGTGCGCC TATCGTAGTC AAGGCGGATG GCTTGGCACT TGGGAAGGGT 19080

GTCGTCGTTG CTGAGACGGT TGAGCAAGCG GTCGAAGCCG CTCATGAGAT GCTTTTGGAC 19140

AATAAATTTG GTGACTCAGG TGCGCGCGTG GTTATTGAGG AATTCCTTGA AGGAGAGGAA 19200

TTTTCACTCT TTGCCTTTGT CAATGGTGAT AAGTTCTACA TCATGCCAAC GGCTCAGGAC 19260

CACAAACGTG CCTATGATGG CGACAAAGGG CCTAACACGG GTGGTATGGG TGCCTATGCG 19320

CCAGTCCCAC ACTTACCACA GAGTGTAGTT GATACAGCGG TTGACACCAT TGTCAAGCCA 19380

GTTCTAGAAG GGGTGATTAA AGAAGGTCGC CCTTATCTGG GAGTTCTTTA CGCAGGGCTT 19440

ATCCTGACAG CTGATGGACC GAAAGTCATT GAGTTCAACG CTCGGTTCGG AGATCCAGAA 19500

ACTCAGATTA TCTTGCCTCG CTTGACCTCT GACTTTGCTC AAAATATCAC AGATATCCTG 19560

GATAGCAAGG AGCCAAATAT CATGTGGACG GACAAGGGTG TGACTCTGGG TGTGGTTGTC 19620

GCATCCAAGG GCTACCCGCT AGACTATGAA AGGGGCGTTG AGTTGCCAGC CAAGACAGAA 19680

GGCGATGTCA TCACCTACTA TGCAGGGGCT AAGTTTGCGG AAAATAGCAG AGCACTGCTC 19740

TCAAACGGCG GACGAGTTTA TATGCTCGTT ACCACAGCAG ATACCGTCAA AGAAGCCCAA 19800

GCCAGCATAT ACCAAGAACT ATACCAACAA AAAATAGAAG GACTCTTCTA CCGAACAGAT 19860

ATCGGAAGCA AGGCAATTAA GTAAAGATAT AAGAATAACG CGCCGTAGTC GCCAAACACG 19920

ATAATGGTCG TCGTGGTGAA AAGACCAGAA CAGTGAATGT TCTGGTCAGG GGGAAACTTG 19980

GAGACCTTAG GCTCAAAGTT TAGGAATGAA ACCGAAGGTT TGCTTCCGCC TCCATCACCT 20040

AAGACCATTA TCAAAAAGAA AAATAAAAAT TCACAAAATA CGTTAATGAT CGTATGGTTT 20100

GCGAGCGTTA GCGAGCTAAT ATAGAACAAT CACCGCCGTT GTGAAAGAAC GATTGGATGA 20160

TAATCCAATC GTTCAGGGAA ATTGGAAGAC CTTGGGTTTC CAATTTAGGC ATGAGACACC 20220

TTTGGTGGCT GCTGCCGTCC CTCACAAGCT AAGGTGATTG TTGAAAAAGA GGAAAAAGGA 20280

GAAGAAATGA AACCAGTAAT TTCCATCATC ATGGGCTCAA AATCCGACTG GGCAACCATG 20340

CAAAAAACAG CAGAAGTCCT AGACCGCTTC GGTGTAGCCT ACGAAAAGAA AGTTGTTTCC 20400

GCACACCGTA CACCAGACCT CATGTTCAAA CATGCAGAAG AAGCCCGTAG TCGTGGCATC 20460

AAGATCATCA TCGCAGGTGC TGGTGGCGCA GCGCATTTGC CAGGCATGGT AGCTGCCAAA 20520

ACAACCCTTC CAGTCATTGG TGTGCCAGTC AAGTCTCGTG CTCTTAGTGG AGTGGATTCA 20580

CTCTATTCTA TCGTTCAGAT GCCGGGTGGG GTGCCTGTTG CGACCATGGC TATCGGTGAA 20640

GCTGGAGCGA CTAACGCAGC TCTCTTTGCC CTCCGTCTCC TCTCTGTAGA AGATAAGTCC 20700

ATTGCGGATG CACTTGCCAA CTTTGCTGAA GAACAAGGAA AAATCGCAGA GGAGTCGTCA 20760

AATGAGCTCA TCTAAAACAA TCGGAATTAT CGGTGGCGGT CAACTGGGTC AGATGATGGC 20820

CATTTCTGCT ATCTACATGG GCCACAAGGT TATCGCGCTG GATCCTGCGG CGGATTGCCC 20880

GGTCTCTCGT GTGGCGGAAA TCATTGTGGC ACCTTATAAC GATGTAGACG CCCTCCGTCA 20940

GTTGGCAGAC CGTTGCGATG TCCTCACTTA TGAGTTTGAA AATGTCGACG CTGACGGTTT 21000

GGATGCCGTT ATCAAGGATG GACAACTCCC TCAAGGAACA GATCTGCTCC GCATTTCGCA 21060

AAATCGTATT TTTGAAAAGG ACTTTTTGTC AAACAAGGCT CAAGTCACTG TGGCACCCTA 21120

CAAGGTCGTG ACTTCTAGCC TAGACTTGGC AGATATCGAC TTGTCGAAAA ACTATGTCCT 21180

CAAGACTGCG ACTGGTGGCT ACGATGGTCA TGGACAAAAG GTTATTCGTT CAGAAGCAGA 21240

CTTGGAAGCA GCCTATGCGC TAGCAGACTC AGCAGACTGC GTCTTGGAAG AATTTGTCAA 21300

CTTTGACCTT GAGATTTCTG TCATCGTGTC AGGAAATGGC AAGGAGGTGA CGTTTTTCCC 21360

AGTTCAGGAA AATATCCACC GCAACAATAT CCTGTCTAAG ACCATCGTAC CAGCCCGCAT 21420

TTCTGAAAGT CTAGTAGACA AGGCTAAAGC TATGGCAGTG CGAATCGCAG AACAACTCAA 21480

CTTGTCTGGA ACTCTCTGTG TGGAAATGTT TGCGACAGCT GATGACATCA TTGTCAATGA 21540

AATCGCCCCA CGACCACATA ACTCTGGGCA CTATTCTATT GAAGCCTGTG ATTTCTCTCA 21600

GTTTGACACC CATATTCTGG GTGTTCTCGG AGCACCATTA CCAGTCATCA AACTCCATGC 21660

GCCAGCCGTT ATGCTTAATG TCCTCGGTCA GCATGTCGAG GCTGCTGAAA AATATGTCAC 21720

AGAAAATCCA AGCGCCCACC TCCACATGTA TGGTAAAATA GAAGCAAAGC ATAATCGTAA 21780

GATGGGACAT GTGACTTTGT TTAGTGATGT GCCGGATAGT GTGGAAGAGT TTGGGGAAGG 21840

GATTGATTTT TAGGACAAGT CTATGATACA AATTATCGTT AATACATTTA TTGAAAAGTA 21900

TAAGACTGGA GCAGTTGTTG AAGTGTTGTA TGCCAGTGCT GACCAAGATA AGGTACAAGC 21960

TAAATATGAA GAACTAGCTG CACAATACCC CGAAAATTAT TTAGCTATCT ATAATGTACC 22020

GCTGGATACG GATTTGAATA CACTAGATCA TTACCCGTCT GTGTTTATTG GAAAAGAGGA 22080

GTTTGAGTAG AAATCTTGGT TTACCTAGAT AGCTTATTCC CAACAGCTTA AGAAGAAAGG 22140

AAAAATTAAC ACATGATCAA CCGTTACTCT CGCCCTGAGA TGGCGAATAT TTGGAGTGAA 22200

GAAAATAAAT ACCGTGCTTG GCTTGAGGTG GAAATCCTCT CTGACGAGGC ATGGGCTGAG 22260

TTGGGGGAAA TCCCTAAGGA AGATGTGGCT TTGATTCGCA AGAAGGCGGA CTTTGACATC 22320

GACCGTATTT TGGAAATTGA GCAGGAGACG CGCCACGATG TGGTGGCTTT CACGCGTGCG 22380

GTTTCTGAGA CTCTTGGTGA AGAGCGCAAG TGGGTTCACT ATGGGTTAAC TTCTACTGAC 22440

GTGGTGGATA CTGCTTATGG TTACCTCTAC AAGCAGGCCA ACGACATCAT CCGTCGTGAC 22500

CTTGAAAACT TCACTAATAT CATCGCTGAC AAGGCCAAGG AGCACAAGTT CACCATCATG 22560

ATGGGGCGTA CTCATGGTGT GCACGCTGAG CCGACAACCT TTGGTCTTAA ATTAGCAACT 22620

TGGTACAGCG AAATGAAACG CAATATCGAG CGCTTCGAGC ATGCGGCTGC TGGTGTAGAA 22680

GCTGGTAAGA TTTCTGGTGC GGTTGGGAAC TTTGCCAATA TCCCACCATT TGTAGAGGAG 22740

TATGTCTGCG ATAAACTTGG CATCCGTGCC CAAGAAATCT CTACACAAGT CCTTCCTCGT 22800

GACCTTCACG CTGAGTACTT TGCGGTTCTT GCCAGCATTG CGACTTCAAT CGAACGTATG 22860

GCGACTGAGA TTCGTGGTCT ACAAAAATCT GAGCAACGCG AAGTAGAAGA GTTCTTTGCT 22920

AAAGGGCAAA AAGGGTCTTC AGCAATGCCT CACAAACGCA ACCCAATCGG TTCTGAAAAT 22980

ATGACTGGTC TGGCGCGTGT CATTCGTGGT CACATGATTA CGGCTTATGA AAACGTCGCT 23040

CTCTGGCATG AACGCGATAT TTCTCACTCA TCAGCTGAGC GTATCATCAC ACCAGATACG 23100

ACCATTTTGA TTGACTACAT GCTCAACCGT TTTGGAAATA TCGTCAAGAA CTTGACAGTC 23160

TTCCCAGAAA ATATGATCCG AAACATGAAC TCGACTTTTG GTCTTATCTT TAGCCAACGG 23220

GCTATGTTGA CATTGATTGA AAAAGGCATG ACCCGTGAGC AAGCCTATGA CTTGGTGCAA 23280

CAAAAACAGC CTACTCTTGG GACAACCAAG TAGACTTTAA ACCACTTCTT GAGGCAGATT 23340

CAGAAGTAAC ATCACGTCTC ACACAAGAAG AAATCGATGA AATCTTCAAC CCAGTTTATT 23400

ACACCAAACG AGTGGATGAT ATCTTTGAAC GTCTTGGACT AGGTGATTAA TTAAAAAATA 23460

AACAGCGAGC TTCAATCTCG CTGTTTATTT TTTATCGAAA AGACTTAGTC TTCTTTTCTT 23520

TTAGTGAGTC CATAGGCTGC TAGTGTGGAC ATGAGTCCTG CGACTACTAG TCCTGCAGAA 23580

TCGTGAGTTC CTGTTTCAGG AAGTTTTTTC TCTGTTACCA CAGGAGCTGG ATCTTGAGGA 23640

AGAACTTTGC TTTCCTCAGC AGGAGCAGTT GATGGAGCTG GTTGGCTTGG GATTTCTAGT 23700

TTTGGTTTTT CTTCAGCAAT AGCGGCTTGT CCGTTTTCAT CGCCTACATG TGTTACCATA 23760

GTTCCGACTT CGACTATTTG AGTAACGGCT TCCTGTGCTA CGACACTATT TACAAGTGTT 23820

TTCACTTCCT TACCATCGGC AGAAGTGCTC ACAGAGTAGA AGTTGCTACG ATGTCCATTG 23880

ACGCCCTTAG TAATGACTTG TGTTTTTCCT TTGAGTAAGA GTGGATTTTC ACAAGTCACT 23940

GTGGTAAATG GAATTTCTTC TTCTTGGATA TCCAGTCTAG GTTTTACCTC AGTAGTTGGT 24000

GCAAGACCAC TTTCATCACC CTTGTGAGTT ACAGGAGCGC CAACTTCAAC CACTTGGTTT 24060

ATAACTTCTT TGGTTACCTG GCTATCAAGG ACTGTTTCTG TTGTTTTTCC ATTTTCAGTG 24120

AGTACAGAGA TGTAATGAGT TCGTTCACCT TTGACTCCTG CTGTGATAAT ATTTTCCTGA 24180

CCGGCTGGGA GGTTAGGATT TTCTTTCTTG ATAACTTCAA ATGGAATTTC TTCAGTTCTT 24240

GTGATGAGTT CTGGTCTGGT TTCAACATTG GCAGCCACTT CATTTTCATC TAGGCTTCCT 24300

GAATGAGTTA CAGCTGGTTT GAGGCCTTGA AGAGCGGCTT TTAGGTTGGC TACAAGCGTG 24360

TCAAGCTCAG CTTGTTTATT ACGGTTGAGG TTGTAATTTA GAGCTGTTTT AGCTGCGTCA 24420

AGGGCCTCAA GACTTTCTTT ACTATATCCT TCTAAGTTTG TAGGAATTTT AGCTAATTCT 24480

TCGCGGAGAG CATTATAATT AGCACGAAAG TAGTCTTTGT TGTGGTCTGC AAAGGCAGTC 24540

ATGAGTTCAA AGATTTCCTC TTCCTTGTAT TCAGCGCTTG GTCTATCTGC CCAGATTGAA 24600

AGCATACTTC CGACTGTTGG AAGATCTACT TCAGGATATT TGGTAGAAGC TAGTTGATTG 24660

AATGGTGTTT TTCCAGTATT CTCAATAGCT TTCTTGAGGA AACCACCACC ATCTTCTGGT 24720

TTTTGACCAA GAATGTAGTA CCAGTCACCG TTGGTATTCA AGAATTTATA GCCTTTGCTT 24780

GCTAGGTATT GAGGTGATGC GAGGTTATAT CCCCACCAGC CTTTAGACCA GTAAGAAATC 24840

AAGACATCTT TGTCAAACTG AACATCGTCC TTGTCTTCAT AGTAGAAGCC ATCGTTGAAG 24900

GCCATTGGTT GAAGCCCTCT TTCTTTGGCC ATAGCTGCGA GGGTGTTGGC ATATTCGGCA 24960

AATTTGCCAT AGAGTTGATA CCACTTGAGG TAGTACCAGC CTTGGGCACT AGTCGCATCG 25020

TTGGCGTATT CGTCAGTACC AAAGTTGAAA ATCTTTGTTT TACCTGCAAA GAAGTCCATG 25080

TATTTACCGA TGAGGGCTTT TACAAAGTTC ATCGCTTCTT CGTTTTTCAA GTCCATAGTT 25140

GTTTTTGAAA CTTTATCAAA GTGGGCTTGA GGATTTTTAA TACCTAATTT TTCCATGGCA 25200

ACCAGCATAG CATCCATGTG ACCTGGACTG TTAATAGCTG GGATGAGACC GATGTCCTTA 25260

GATTTAGCGT ATTCAATTAG CTCTGTTACT TCTGCCTGTG TTAGTGCAGT ACCGTTTGGA 25320

TCGTCGTAGT AAGCTTTAGT TCCTTCGATA ATAGCTTTTT TAACGTCATC ACTAGCATAG 25380

GTTTTTCCGT TGGCAGTAAT GGTCATATCA TCGAGTAGAA AGCGAAGTCC GTCATTTCCT 25440

AGAAGGAGAT GGACATCAGA ATATCCGAGC TCACTGGCCT TGTCTACGAT GCGTTTGAGC 25500

TGGTTCAGAG TAAAGTATTT GCGTCCAGCA TCGATTGAGA TTACCTTGTT TTTGGCAAGT 25560

TTTTCAACCT CACGTTTAGC TTCTTCTTCT TTTTGAGCTT CAGGCGTGAG GGTCAAGTTG 25620

TTGACAGTTT CTTGAAGTTT AGCAATGGCT TGATCAATCG TATCTTGTTG GGCACGGCTA 25680

AGGTTGCTAT CGAGAGAGCG AATAGCTTTT TCAGCTTCTT TTACGGCCGT GACGCTTTCT 25740

GCAGTATAAC GGTTCAGGTC TTTTGGTACC TCGTTAAGTG CTTGCTCTGC AGATTCATAA 25800

TCAGCTGCGA AGTATTCAGC GTTGGCATTT GCAAAATGAC GCATGAGTTT GAAGAGGCGT 25860

GATGGTGAAT AACGTGCAGA TGGAGTGTCA GCCCAAGCAG CTACCATACC ACCGATGATT 25920

GGGATATCAG CTCCTTCTGT TTTTGGTACA GAAGTGATTG GTGTGTTTTT AATACCATTG 25980

AGCCCCTGAT CGAGATTGTA CCAGCCTTGG CCATCAGCGT TTCGTCCAAG AACGTAGTAC 26040

CAAGCATCAT TGGTATTAAG GATTTGGTGA CCTTTTTCAG CTAGTAGTTT AGAAGAAGCG 26100

ACATCGTAGC CTCCCCAACC ACCAGTCCAC ATAGAAACGA TGATGTCTTT GTCAAAACTA 26160

CCAAAGCTTG TGTCGCTATT GTAGTAGATA CCGTCGTTAA AAGCCATTGG TTTGAGACCG 26220

TGCGATTTTA CAATACGAGC GAGGTCATTG GCGTAGGCAA TAAATTTTTC ATAGCCTTTT 26280

ACAGGGTAGC CTTCGTTTGG ATAGTATTTA TCAGCTTGAA GCACACTCCA ACCTTTAGCA 26340

TCTGTCGCAT CATTGGCATA TTCATCAAGT CCGATGTTGA AGATTTCAGT CTTTTTCGCG 26400

AAATAAGCAG CATACTTGTC GATAAGGGCT TTTGTAAAAG CGACAGCTTG TTCGTTGTCA 26460

AGATCGACAG TACGGGCTGA TTTCTTCCCA AAATAGCTAA AGTTAGGGTT TTGGATTCCC 26520

AATTCTTTCA TGGCATTGAG AATCGCATCC ATGTGTCCAG GACTATTTAC TGTCGGAATG 26580

AGACCGATAC CTTTATCTTT GGCATAGTTA ATCAGATCTG TCATTTGACT TTCTGTTAAG 26640

TGATTGCCGT TTGGATCGTT GTAATAATCA TTTGTACCTT TTTCAATGGC GCGTTTGACA 26700

TCGTCACTGG CATAGGTCTT GCCGTTAGCT GTGATGCTCA TATCGTCCAA CATGAAACGG 26760

AGTCCATCAT TTCCGACTAA TAGGTGTAAA TCAGTGTAGC CATAATGTTT CGCTTTATCG 26820

ATGATTTCCT TGAGCTGTTC TGGTGAGAAA TATTTACGTC CAGCATCAAT AGAAACAATT 26880

TTCTTTTTCG CTAGTTTTTC ATTTACAGTT GCAGCACGTT CCTTTCCTGC CTCTGTTGCC 26940

GGTTTGTCAG CCTCTGCTTT CGCTTCATCT TTTTTAGCTG GTTTATCCTT GTCAGTCTTG 27000

TCTGTATTTG ACTCTTTAGA ATCAACCTCT TTCGCTTCTT CCTTTTTAGG GCTAGCTTCT 27060

TCTGCCTTTT TATTAGCAGT TTCTTTTTCA GCAGAAGTTG GAGTTACCAC TTCTGCTTTA 27120

TCACTAGGAG TTGAACTAAC TTCCTCTTGT GGTTTTTCTT CTGTTTTTGG AAGACTAGCT 27180

ACCTTATCAG TAGCTGGAGT TTCTGTTTCT ACAGTTTTTG GAGCTTCTGG TTGAAGCACT 27240

GCTTTAGGTG TTTCCTCAGT CCGATTTTCG GATGATTGAG GGGAATCAGA AACCGTATGG 27300

ATGGTCGGTT GGTTTTCTGT AGTAGTAGGA GTAACTCCAT CGGCTGCAAC AGTCTGTGCT 27360

TGGAAGGCAA ATCCAATTAG AACAGAAGCT GCTCCTACAG CGTATTTACG AATAGAAAAA 27420

CGCTGTTGTT TTTCATGTTT CATTGCAAAA CCTCCTGATT GCATTGTTAT ATTGATAGCG 27480

ATTATATAAA TCAACGCCTT TATTTTATTT CTTATATTAA TTTCTTATAT TAACGAGAGT 27540

CAAGAGGAGA TGACAAAAAA CTATAATAAG TATAAAAAAA TATAAAATTT AAACTTAAGA 27600

TTTCAGATTG GTCGGAAAAA ATACGTATAT ATATCTAGTA TAATTTTTGG TTCTATTTCT 27660

ATAAAATATT CCACAAATTA TAGAATTTTC CAAAAATAGG TAAGCGCTAC CTTTTTGGTG 27720

TAGTATAATA AGCATAGAAA AAGCCCAAGC GATTAGCTCA GGTTTTCTTC TTAGTGATCA 27780

CGGTCACATG AGATAAATTT AATCTTGTAG TAATCAGATC GTTTGTAAGT TTCACTGTAT 27840

TCTAAAACTT GGCCAGTTGA TTCGAGTTTG GTGATTTTAG TTTGTAGGAC AGTAGGGAAT 27900

TGTTCATCGA CTCCGAGGAC TGAAGCTGCA TGTTCTGGAG TTGGAAAGAC TATTTCGTTG 27960

ATTTCTTCAA AGTGTTCATC ATTCATGTGA ATGTGGTAGT CTAACTTGAA ACGATTATAG 28020

ATAGAACTAT AGTATTCAAG GTTTGGATAA TTTGCGTTGA TATATTGTTC TGGGATGTAG 28080

GATGTATGGT AGATATAAAC GACACCGTTT GATTCGCGGA TACGTTCAAT CTTGTAGTAG 28140

AATTGATCGC CGCGTAGACC CAATTTTTCC AAGTAAACAA GCTTGTTTCC GCGTTCAATT 28200

GAAAGAACAG TTACCTTATC ATCTTTAGCA TTGAAGAGTT CAATATCTGA AAACTCTACA 28260

AGCTTGTGTT TGCGTGCACG TGAAACGAAG GTTCCTTTTC CTTGTTGGCG GACAATATAG 28320

CCATCTTTGG CAAGGTCGTT TAAGGCGCGA ACAACTGTGA TAGAGCTGAC ATCGTACATT 28380

GAAATGAGTT CTGCTTCAGT GTAAAATTTA TCTCCACTGC TAAACTGCCC AGAGATGATT 28440

TTATTTTTTA ATTCGTCTTT TATGTATTGA TGG 28473

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6749 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

CCTGATGGGT GGTATGCGAG GATACAGTTC TGAAAATCGC CGTTACTTAA TTAATGGACG 60

CGAAGTCACA CCTGAGGAAT TTGCTCACTA TCGTGCGACT GGTCAATTAC CAGGAAATGC 120

AGAAACTGAT GTGCAAATGC CACAACAGGC ATCAGGTATG AAACAAGGCG GTGTCCTTGC 180

AAAACTAGGT CGAAACTTAA CAGCAGAAGC GCGTGAGGGC AAGTTGGATC CTGTTATCGG 240

ACGAAACAAG GAAATTCAAG AAACATCTGA AATCCTCTCA CGCCGCACCA AGAACAATCC 300

TGTTTTGGTC GGAGATGCAG GTGTTGGTAA GACAGCAGTT GTCGAAGGTC TAGCGCAAGC 360

CATTGTGAAC GGAGATGTTC CTGCTGCTAT CAAGAACAAG GAAATTATTT CTATTGATAT 420

CTCAGGTCTT GAGGCTGGTA CTCAATACCG TGGTAGCTTT GAAGAAAATG TCCAAAACTT 480

AGTCAATGAA GTGAAAGAAG CAGGGAATAT TATCCTCTTC TTTGATGAAA TTCACCAAAT 540

TCTTGGTGCT GGTAGCACTG GTGGAGACAG TGGTTCTAAA GGACTTGCGG ATATTCTCAA 600

GCCAGCTCTC TCTCGTGGAG AATTGACAGT GATTGGGGCA ACAACTCAAG ACGAATACCG 660

TAACACCATC TTGAAGAATG CTGCTCTTGC TCGTCGTTTC AACGAAGTGA AGGTCAATGC 720

TCCTTCGGCA GAGAATACTT TTAAAATTCT TCAAGGAATT CGTGACCTCT ATCAACAACA 780

CCACAATGTC ATCTTGCCAG ACGAAGTCTT GAAAGCAGCG GTGGATTATT CTGTTCAATA 840

CATTCCTCAA CGTAGCTTGC CAGATAAGGC TATTGACCTT GTCGATGTAA CGGCTGCTCA 900

CTTGGCGGCT CAACATCCAG TAACAGATGT GCATGCTGTT GAACGAGAAA TCGAAACGGA 960

AAAAGACAAG CAAGAAAAAG CAGTTGAAGC AGAAGATTTT GAAGCAGCTC TAAACTATAA 1020

AACACGCATT GCAGAATTGG AAAGGAAAAT CGAAAACCAC ACAGAAGATA TGAAAGTGAC 1080

TGCAAGTGTC AACGATGTGG CTGAATCTGT GGAACGAATG ACAGGTATCC CAGTATCGCA 1140

AATGGAAGCT TCAGATATCG AACGTTTGAA AGATATGGCT CATCGCTTGC AAGACAAGGT 1200

GATTGGTCAA GATAAGGCCG TAGAAGTTGT AGCTCGTGCT ATCCGTCGTA ACCGTGCTGG 1260

TTTTGATGAA GGAAATCGCC CAATCGGCAA CTTCCTCTTT GTAGGGTCTA CTGGGGTTGG 1320

TAAGACGGAG CTTGCTAAGC AATTGGCACT CGATATGTTT GGAACCCAGG ATGCGATTAT 1380

CCGTTTAGAT ATGTCTGAAT ACAGTGACCG CACAGCTGTT TCTAAGCTAA TTGGTACAAC 1440

AGCAGGCTAT GTGGGTTATG ATGACAATAG CAATACCTTA ACAGAACGTG TTCGTCGCAA 1500

TCCATACTCT ATCATTCTCT TGGATGAAAT TGAAAAGGCT GACCCTCAAG TTATTACCCT 1560

TCTCCTCCAA GTTCTAGATG ATGGTCGTTT GACAGATGGT CAAGGAAATA CAGTAAACTT 1620

CAAGAACACT GTCATTATTG CGACCTCAAA TGCTGGATTT GGCTATGAAG CCAACTTGAC 1680

AGAAGATGCG GATAAACCAG AATTGATGGA CCGTTTGAAA CCCTTCTTCC GTCCAGAATT 1740

CCTCAACCGC TTTAATGCAG TCATCGAGTT CTCACACTTG ACTAAGGAAG ACCTTTCTAA 1800

GATTGTAGAT TTGATGTTGG CTGAAGTTAA CCAAACCTTG GCTAAGAAAG ACATTGACTT 1860

GGTAGTCAGT CAAGCGGCTA AAGATTATAT CACAGAAGAA GGTTACGACG AAGTCATGGG 1920

GGTTCGTCCT CTCCGTCGCG TGGTTGAACA AGAAATTCGT GATAAGGTGA CAGACTTCCA 1980

CTTGGATCAT TTAGATGCTA AACATCTGGA AGCAGATATG GAAGATGGCG TTTTGGTTAT 2040

TCGTGAGAAA GTCTAAGACA GAATTTTGAG GATAAAAAAG AAGGAGCCAG CTGAAAAAAA 2100

CTGGTTCCTT TTTAGGTACG ACAGGCATGT CGTATAGTAG AAGTGTATTA TTCTAGTTTC 2160

AATATACTAT AGTAGCTCAG AAGTCGGTAC TTAAACGTGC TATATCAAAA CCAGTCCTGG 2220

AAAAACGTGG ACTGGTTTCG TGTTTGGATT ATTACCTTGA ACGACATGCG TTAAAAGTTA 2280

GTTGAACCGC CGTATGCCGA ATGGTACGTA CGGTGGTGTG AGAGGGGCTA GAGATTATCC 2340

CCTACTCGAT TTTAAATCAC ATGACGTTCA AAGGCATCAT CTGAAATCCC TTGTTCCAAG 2400

ATGAGTTTTG CCCATTCTTT AGCAGAGAAG AGGCTGTGGT CCTTGTAGTT TCCGCAAGAT 2460

TCGATGGTTG TCCCTGGGAC ATCTTCCCAA GTAGTAGTTT CAGCGATTTC CTTGAGCGAA 2520

TCCTTGATAA CAGCTGCGAT TTTAGCACTG GTGTGACGTC CCCACATAAT CATGTGGAAG 2580

CCTGTGCGGC AACCAAATGG TGAACAGTCA ATCATGCCGT CAATGCGGGT ACGGATGAGT 2640

TTGGCTAAGA GGTGCTCGAT AGTGTGAAGG CCGGCAGTAG GGATAGAGTC TTCGTTTGGT 2700

TGCACCAAGC GAATATCATA ATTGGAGATG ATGTCTCCTT TTGGTCCTGT TTCTTCCCCA 2760

ATCAAGCGAA CATAGGGTGC TTTGACAATG GTGTGGTCAA GTTCAAAACT TTCGACAATA 2820

ACTTCTTTTG ACATGGTAAA TCCTTTCAGT TTTCTTCTCT CATTATATCA TAAAGGTTGC 2880

TCCTGAGACA GAGAGAAAAC CTCTCCGAGG CTGGAGAGGT TGAAATCTTT ACTTACGATA 2940

TAAGCGGTCG TATTGGTAGT ATGGGTCAAA GGTTACGTTG ATACCCAGTT TACGAAGGAC 3000

ATTCTTGTCT TCATCAGTCA AGATGATGGT TGAGTGGGCT TCGCTTCCTT TGAGGTTGCC 3060

GAGTTCTTCC ATAGCGCGGG CAGCATCAGG ATTTTCTGTA GCTGTGATAG CAAGTGCAAT 3120

CAGGATTTCA TTTGAATGAA GGCGTGGATT GCGGCTACCG AGATGATCGA TTTTAAGACC 3180

TTGGATTGGC TTAACAACTT CAGGCTCGAT TAGTTTTACT TCTTTAGCGA TGTCAGCTGA 3240

TTTTTTGATG GCGTTGATCA AGGCAGCGGC TGTAGGACCA AAGAGTTCTG AGTTCTTACC 3300

AGTGATGATT TCCCCATTTG GCAATTCAAA GGCTAGGGCT GGTCCACCAG TTTCTTCTGC 3360

TTTTTGGCGC GCAACGACAG CAACCTTACG GTCTGCAGGT GTGATACCGA GGTCGTTCAT 3420

GAGCAACTCA ATTTTCTTGA CGGCAGCTTC GCCAACTTTT TCAGCTTTGA AGTCAAGAAC 3480

TGTTTGATAG TAACGGCGGA TGATTTCTTG TTTAGAAGCT TCGACAGCGG CCTCGTCATC 3540

TGTAATAGCG AAACCAACCA TGTTGACACC CATATCTGTC GGTGAAGCGT ATGGTGATTT 3600

TCCGAGAATA CGTTCCAACA TGCGTTTGAG CACTGGGAAG ATTTCGATAT CACGGTTGTA 3660

GTTGACAGTG GTTTCTCCAT AGGTTTGAAG ATGGAAGGGG TCAATCATGT TGACATCATC 3720

AAGGTCAGCT GTGGCAGCTT CATAAGCCAA GTTAACTGGA TGATGAAGGG GAAGATTCCA 3780

AACAGGGAAG GTTTCAAATT TAGCGTAGCC AGATTTGATG CCATTGATTT GGTCGTGGTA 3840

CATATTGGAC ATACACGTTG CCAATTTTCC AGAACCAGGT CCAGGAGCGG TTACGACAAT 3900

CAAGTTGCGA CTGGTTTTGA TGTAGTCGTT TTTGCCCATG CCTTCTGGGG AAATGATGTG 3960

ATCCATATCC GTCGGATATC CTTTGATTGG ATAATGAAGA TAAGAATCAA TTCCGTTTTT 4020

CTCAAGTTGA TTGCGGAAGG CATCTGCAGC GGGTTGGCCA GCGTATTGTG TAATGACAAC 4080

GGAACCAACA AAAATCCCTA ATTCATTGAA TTTATCAATC AAACGAAGAA CTTCTTGGTC 4140

ATAAGAAATG CCTAAGTCGC CACGTGCTTT GGAATGTTCA ATGTTGCTAG CATTAATGGC 4200

AATCACAACC TCAACCTGCT CTTTCAATTC TTGCAAGAGC TTGATTTTGT TGTCAGGTTC 4260

ATAACCAGGA AGGACACGAG CAGCGTGGAA ATCTTCTAAC ATTTTACCGC CAAACTCTAA 4320

GTAGAGCTTG CCGTCAAATT GGTTAATGCG CTCCAAAATA TGGTCGCGTT GTAAATTCAA 4380

ATATTGTTCA GAACTAAAAG CTTGTTTTTT CATTTTTTTA CCTCTGGACT CTATTATAAT 4440

AAAAAATTGG AAGTTAGGAA ACTACGGAGC TAAAAAAGAA ATTAAAAAGA TTAAGCAAAC 4500

GCTTGCACAA AATTTTAAAA AGTGCTATCA TAGACTATAG ATTATGAAAA TAATGAGGTA 4560

AACAGATGCA AGAAAAATGG TGGCACAATG CCGTAGTCTA TCAAGTCTAT CCAAAGAGTT 4620

TTATGGATAG TAATGGAGAT GGAGTTGGTG ATTTGCCAGG TATTACCAGT AAGTTGGACT 4680

ATCTAGCTAA GCTAGGAATC ACAGCAATTT GGCTTTCTCC CGTTTATGAC AGCCCTATGG 4740

ATGATAATGG CTATGATATT GCTGATTATC AAGCGATTGC GGCTATTTTT GGAACCATGG 4800

680

AGGACATGGA TCAGCTGATT GCAGAAGCTA AGAAGCGTGA CATTCGTATC ATCATGGACT 4860

TGGTGGTCAA TCATACCTCA GATGAACATG CTTGGTTTGT CGAAGCCTGT GAAAATACTG 4920

ACAGCCCTGA GCGAGACTAC TATATCTGGC GCGATGAACC CAATGACCTA GATTCTATCT 4980

TTAGTGGGTC TGCTTGGGAA TACGATGAAA AGTCAGGTCA ATACTATCTC CACTTTTTCA 5040

GCAAGAAACA GCCGGATCTC AACTGGGAAA ATGAAAAACT TCGCCAGAAA ATTTATGAGA 5100

TGATGAACTT CTGGATTGAT AAAGGTATTG GTGGTTTCCG TATGGATGTT ATTGACATGA 5160

TTGGCAAAAT TCCTGACGAG AAGGTAGTCA ATAATGGTCC TATGCTCCAT CCCTATCTCA 5220

AGGAAATGAA TCAGGCGACC TTTGGAGATA AGGATCTCTT GACAGTAGGG GAGACTTGGG 5280

GAGCAACTCC AGAGATTGCC AAGTTCTACT CTGATCCAAA GGGGCAAGAA TTGTCTATGG 5340

TCTTCCAGTT TGAACATATC GGTCTTCAGT ATCAGGAAGG TCAGCCTAAA TGGCACTATC 5400

AAAAAGAGCT GAATATCGCT AAGTTAAAAG AAATCTTCAA CAAATGGCAG ACAGAGTTAG 5460

GAGTTGAGGA CGGCTGGAAT TCCCTCTTCT GGAACAACCA TGACCTCCCT CGTATTGTCT 5520

CAATCTGGGG AAATGACCAA GAATACCGCG AAAAATCTGC CAAAGCCTTT GCAATCTTAC 5580

TTCATCTCAT GAGAGGAACT CCTTATATCT ACCAAGGTGA GGAGATTGGG ATGACCAACT 5640

ATCCGTTTGA AACACTGGAT CAAGTAGAAG ATATTGAATC TCTCAACTAT GCGCGTGAGG 5700

CTCTTGAAAA AGGTGTTCCG ATTGAAGAAA TCATGGACAG TATCCGTGTT ATTGGACGTG 5760

ACAATGCCCG TACCCCTATG CAATGGGACG AGAGCAAAAA CGCTGGTTTC TCAACAGGTC 5820

AACCTTGGTT GGCGGTTAAT CCAAATTACG AGATGATCAA TGTCCAAGAA GCGCTGGCAA 5880

ATCCAGATTC TATTTTCTAT ACCTATCAGA AACTGGTCCA AATTCGCAAG GAGAATAGCT 5940

GGCTAGTTCG AGCTGACTTT GAATTGCTTG ATACGGCTGA TAAGGTCTTT GCTTATATAC 6000

GTAAGGATGG CGACCGTCGC TTCCTAGTTG TGGCTAACTT GTCCAATGAA GAGCAAGACT 6060

TGACAGTAGA AGGAAAAGTC AAATCTGTCT TGATTGAAAA CACTGCGGCT AAAGAAGTAC 6120

TTGAAAAACA GGTCTTGGCT CCATGGGATG CTTTCTGTGT GGAATTACTA TAAATATTTT 6180

TTGCAGAAAA ATTTAAAATT GAAATCGTAT AAAAACAAGG GAGGACTGTA TAAAAGACAG 6240

AAATCCTTTG TTTTTTATAA CCAAAGTTTA TAAACTTTCA TTCTTGAAAT TCAATTAACT 6300

TTACAAATTC CCACTATTAA GGAGAAAGAA GATGAACATA AAGAAGCGTG TCCTTAGTGC 6360

AGGCCTGACT TTTGCATCTG CTTTGCTTTT ACCCAAATCA TTCATACCTC TCTCAACTAG 6420

ATGTAACTTA CAAAACCCCT GACCTCATGA GCCACTTTCT TCCTCCTCAT GAGGTCAGTT 6480

TTACTTTCTG CTGTTCCAGT ATCGTTTTTC CTCGCTAGAT TTCCTCAAAA GGGCAGACTC 6540

CTCCCTTGGT GCGTCACACG ATTTTTTCAT CTCGACTGTT CTTTAATGCA TCATTAACGA 6600

CGCTTTTCTT CTAGGTGGTT CATAAGGAAC AGGAAGATTC AGGTTGACTT TTCTAATCCT 6660

AGAATAAAGT GCTGAAAACA ATTCGGAATA GGCATAGAGA CTAGACAATT TGAGGAGCTG 6720

CTTGCGTCCT GTTCGAACAC ATTTTCCGG 6749

(2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1842 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

TCTACCCATG GACTTTGAGG CATTCATTGT TCCATCTTCT AGTGGCGAAT CTTTTGATAC 60

AAACGATTCA ATTCACTTGG ATAGTGAAAC TCTCCCGCAA ACATTTTTCT GGTTAACTCA 120

ATCCAGCTGA TATTTCTTTC AGCCAAAATA ATGGACAAGT TCTCCCAAAA TCGTTCAGCC 180

ATATTGCTTC TCCTTTAGTT AGATAAATAA TGTGTTTGCG CCATGTAAAT CAATTGTTTC 240

GTATCTCTTG GCAATAGAGC TCTAGCCTCT TCCAAATTCA GACTTGGATA AACTCGCTTA 300

TTTGAAACCG CAAGAGGAAG TCTGATGGTT AGTTCAGGAT TTTTTAAAAT TATCTCAACG 360

AAATCCGTTA ATCTTAGATT GTCACGGTTC TTAAATCGTA ATAAATTGGG AGATAAAAAC 420

TCAAAACAAT CTGAAGAATA GCTCATCATC TCAATTAATT TGTCCTTTGT CATTTCAGAA 480

ACTGAATGAC AAGATACCTC TATGCCATAG TTTTGGAAGA AATCTAAAAG AAGTTGATTT 540

CTTTGTCTAT TTTTACTTAG ATAGAGATCA ATCATGGGAG ACCTCCCAAA GATTCGGTTC 600

CATTTGATAT TCTGACACGA TTAAGGAATC TAATAAATTA AGGAATCTAA TAAATTTGCG 660

AAGTTAATCG GTTTCTTGTC TTCATCATAA GCTTTTACAG TTACTTGGGT TGTAAGTATT 720

CCCTCTTTTC CCTCGGCTCG ATAGCCTTGT CCATATAAAA CAAAAACGAG ATTTTGATGA 780

TCATCTACAA AGGCATCAAC CCCATTCTTT ATGTCTTGAC TTTCAAGGAA TTCCATAACG 840

TTTTGAAGAT AGGATTCGTA AAATAGTGGG TAGTTATGTT TTTTATGGTA ATCATCTAAA 900

AATGTCACTT CAAACTCACA TGGAGAGTAA TTTTGACTTT GAACAGCCTA AAAGTGCCAT 960

CAAATTTGAA TTGGAATAAA TCAAATAAAT AGCCCCATCC TCATCAATCC AACCTTTGCT 1020

CAAAGACAAC TCCAACCGAT CTTTTAAAAC TGAGTAAACC ACCTTAACCT CCAGTTTCAT 1080

ATTCTTATAC CGTTCACTCT CAAATAAAAG TTTGGGGAGC TTATAATAAC GCTCTGATGT 1140

CTGATATTGA TTAGCGGTAA TACGCTTCAT TATTGTCCCT CCAAGACTAA AATTCCAACA 1200

TTTCCAAATT CATCAAATCG GATTAAACCT ACTTGTTCCA TTTCATCAAC TAACTGAGTT 1260

GCTTTTACCC AAATCATTCA TACCTCTCTC AACTAGATGT AACTTACAAA ACCCCTGACC 1320

TCATGAGCCA CTTTCTTCCT CCTCATGAGG TCAGTTTTAC TTTCTGCTGT TCCAGTATCG 1380

TTTTTCCTCG CTAGATTTCC TCAAAAGGGC AGACTCCTCC CTTGGTGCGT CACACGATTT 1440

TTTCATCTCG ACTGTTCTTT AATGCATCAT TAACGACGCT TTTCTTCTAG GTGGTTCATA 1500

AGGAACAGGA AGATTCAGGT TGACTTTTCT AATCCTAGAA TAAAGTGCTG AAAACAATTC 1560

GGAATAGGCA TAGAGACTAG ACAATTTGAG GAGCTGCTTG CGTCCTGTTC GAACACATTT 1620

TCCCACCACG TGAAGAAAAA GATGGCGGAA GCGTTTGATT GTTAAAGTTT GGAAGTCACC 1680

TCCAGCTAGA TGTTTGAGAA AAAGATAGAG ATTGTAGGCG ATACAGCTCA TCATCATACG 1740

AACTTCGTTT TTGATTAAGG TTGAACTATC CGTTTTATCG CCAAAAAATC CCTCCTTCAT 1800

CTCCTTGATG AAATTCTCGG CTTGACCACG TCCACGATAA AG 1842

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19390 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

TCATCTTTAT CTCCTCGAAA TTTTGTAATA TAGCCATTAT AACAGAATTT TGTGAAAATT 60

CCTATTATAG TAAATCACTA TTTCAGTATA AAAAGAAAAA ACGAATCAGA CGATTCGCTC 120

TTCTTAAAAT CTGAAAATAG CTTTCCAGAA AGGATTAGCC GATTTTTTGC AGATTGAGCA 180

CTGCATCGTG ACTCATCAAG ACTTGACCAT ACTCTTGTAA GACTGAGCGA CTGATATCAC 240

TATCGTCTGC AAACTCGCGC ATACGGGCCA ACAGCCAAGC TGGATATGGG CTTGGATGAT 300

TTTCAATATC CACTAAAATG GTCAAATAAT AGCGCTCGTT CATTTTGTAG AGTTCAGAAG 360

TTTCCATTTC AAAAGTCACT GTCTTGGCAA AAGCTACCAA GTCAGCCAAC TTAGCAAAAG 420

AAAGGATGTA GTAGATGTAA GGTTCTTTCT TACTCTCAGC TTCTTGTTCA GCCTGCTCTT 480

GCTCTTCTTC CTTGACTTCA ACTTGCTCAA GAGATTGAAT GGCTTCGATA TCATCCTTGG 540

TTTTGTCTGC GATGCTTTTT TCCAGGGTTT TGATAAATTC ATCTGGAGAC ATTTGAGCCA 600

ATTCTTCCAT ATCTGGCAAA TCCGATAAGT CTTCAAAATC TAGATTTTGG TCAATCTTTG 660

ACTTGGTCAC AAAGACATCT ACCTTATCAG GTTTTGGAGT CACACGGAAG CTCAACATGC 720

CTGTATCCAG AAAGCTATCA GGCATCTCTA GCTCATCCAA GATAGCATAA AAGAACTCTT 780

CTGTTTTTTC TTGAGGAACG AGAAAGTCAG CAATCTCCAT TCCACGATCC ATCAAATCCT 840

CTAAAGATAT CGTGATTTTT AAAGTTGTAT CACTAATTTG TTTCATTTTC ATTGCTAGTA 900

ACCTCATACT TTCAGTTCTA TCTATTATAC TAGATTTTTA CGATTTTATC AAAAGAAGGC 960

TCCTCTATAC GGATAGATTT TCCCTAGGGT CTTTCTATAG GAGACTCCAA AAGAAAATTT 1020

CTGCAGACAG ATAGAAAAAG CCTTCAAAAT CGGCTAAGAG CCGACTTTGA AGACCTTATA 1080

CATCAGAATA CTTATAATTT AAAGGTTGCT ACACCGAGGA TAGAACGATT TAAGTTTCTG 1140

AGAATTTGAA GACTTTGCTC AAATTTCTTA TAACGAGTCA CTCCGTACTC TTCAACAAGA 1200

AGGACTGTAT CTCTTTCCAA AAGAGATGAT ACATCCTGTA AATCTACAAA ATGCATTCCT 1260

TTTAAAGCTT CTTGACTCTG TTTCAATTTA TCTAAGATAG CTTTATTTGA GCTAACGATG 1320

GTCAATTCCT GTCCAGTATT TTTGTATGAC AAAACATCTG CTAGGTTAGC AATTGTTGTA 1380

ATCTCTGTTA CAAAATCAAT TTGATACTGA GAAAAATCAC CTACTCTATT GATTGTTGGA 1440

TTAAAGAGAT AAACTAACAC ATTTCCCATC ACAACCAAAA TCACACAAAC CACTCCAATA 1500

ACAACTAAAC GAAGAATCAG ATTTTTCACA TTTAAGCCAA GCGCTGTTTC ACCATTTGCG 1560

TTCAATTCTT TAGAGTTGAT GGTTTCCAGT TTTTCAATTT TCACATTTGC ATAGGCATGT 1620

TTAAATTTCT CAATCAACCC ATCAATTTTT TTCTCTAACA AGTTATTGGC ATCTTTACTT 1680

GATGTCAAAA TTTTCACACC AACCCCTGCA TCGTCAATCA TATAGTAGAC GGTCAATTTT 1740

TTCCACCAAT AGTCATTCGT TGAATTTTTC AAGGTTGTTT CTGTCGTGTC TAATTCACTG 1800

GCAATTTTTT TCAACTCACT GGGTTCTACA TCATTGAAAA GATAAGCTCC ATTCAAATTA 1860

CCATCAATCA ATTTCCCATA AAAATCACTA TAACCACCAA TTTGATGATT CAAAATCGTT 1920

TTGTCCGACT CTTTTGGAGG AGTGATTTTA TAGATAAGAT AAGTTGAATA ACTTGTTGTA 1980

TCTTTGACAG TGTTTTTATT CCTAACTGCT TTAATTGTAA ATGGTACAGC AATGAGAGCA 2040

AATAAAGCGA TGAGAGCTAA AATATTTGCT TTTCGCTTTT TATAAAGATT TGCAAACAAA 2100

TCAGCTACTG AATAATGTTC AAACATGATT TTTTTCTCCT TTGTTTAGTA GATACTAGTT 2160

TTCCTTTGTA AGCATTTTTG CTACAAATAT AATCACAAGA ACAATTCCCC AGAATTGCAT 2220

TGTAAATAAA TTGAAGAAAC TTTCTGAAAA GCTGCTTCTT GGCATAAAGA ATAGATTATT 2280

CAAGATGAGT AGGGATAAAG CAAATAGGAT TGTCCTTGAG CGATAGGCTA CTTGCAGCAT 2340

GGCTATAAAT AATACGCCGA GTAAGAAACT AAGCAGAAAG ACTCCAATCA TACCATAGTC 2400

GGTATACAAC TCCATGATAT AACTACTTCC GATACCATGC CCTTTCAAGT ATTCCTTGTT 2460

CAAGACAAGA TAGGATAGAT TGTGGGCATA ACTATTACTA TCAATAGCTA GTTCCACACT 2520

ATTGGTTGTA TGTTCAAAGG CTTTTCCTCC GAAAATGGCT CCCAAACTCC CCCTTGCAAA 2580

ATAATCAAGA ACAGGACCAA AAGTAAAATT ACGGAAATCT CGGTAAGGGA GGCTACTGTT 2640

AAATAGAAAA CCTCGAGCCA GAACACCAAA ACTAGTCCCT TGTTTATAGA TAAAGTCAAG 2700

TAAGATATCC CAGAAACCTG TATGGGAAAC TTGGACATTA TCCCGTACAT AATTGAGTAC 2760

TCCCATCGCT AACATGAGAA TAGGAGAACC TACAAAAATC GCTAACTTTT CTTTAAACCC 2820

AATCCATTTT CCTTTTTCAG TTTGCTCCCG CATAAAGTAA TAAACAAAAG CAAATmAAAT 2880

ACTTAAAATA AAGGGATTTC GTGTCCCAAT TGCCAAATGA ATAGTATTAG CTGCAATAAA 2940

GGAGACAAGC ACTGCTGTGG CCTGCAATTT CTTTGGCTTG GTTGCCAGAT ACATACACAT 3000

TGCATAGACC GTAAAGGTAG ACAAAATGTA GGTAAAATAA GGCAGTTTAC TTTCAAAATT 3060

TGCATAGTAG GCATAGTAGG AAGTCTGCAA ACGATACAAG AGCCGTTCAA ATAACCGAAT 3120

GAAATAGAAA GGATAAGTTA GAAGAAAAAC TCCTAGTGAT ACAAAGCGTA ACCGCTTGAT 3180

ATAAACCTCT TTTAGAGAAT TTCCTATATT TGCTACTTTT ATTTTCTTCC TAGCTATGAA 3240

GTAACGAGCC AGAATGCCTC CTGTGGTCAA GCCCAGAATC GAAATCATGA CAACTATAAA 3300

GGCAAAACGA TAGGCTATTG GATGATAGGT ATCCAAAGCA CCATCCCTAA AATAATCAAT 3360

GGTCGGTCTT GATACCAGAA ATACAAAAAT GGTTAAATAG AAAATAAAAT GGATTAAGTA 3420

ATACTTGATA TCATTCCAAC AAGCAATTAA GCTACTAACC AACAAGAACA ATAAAGTAGA 3480

AAGTAAGCTA ACATTATTAT TATTAAACAG ATACACAATT CCACTTACTA GCGTCAAGGC 3540

ATAACTGACT ATGGTCAAAC TAAATAATAA TCGTTTCCCA TCAATCACTT GGTCACCCCC 3600

GTTCTAATGT AATTTTTTAG ATTTTTCAAT ATTTTTCAGT AATAAGAATC GATATAAGGA 3660

AATATTTATG AATAGGGCCA AAGCACTAAT TCTTCTCCCC TTACGGAAAA TTGGATTCCT 3720

AGAAATAGCA AAGGCATGGC CTTTTAAAAA ACGATGAATC TGAGAATAGG CTTCAAACTG 3780

TTTATACTGA TCATCTAGCA ACATCTTATC CAGAATAAAG AAGTGGGCAT AGGCCAATCT 3840

GAAAAAAGCG ACCTCTTTCA AGTCAGGATA GTTTTTCACA ACTTCATTAT AAAACTTTTG 3900

GTAGATATCA ATATAGGCTA AATCCTTCTC TGCATAGGGT TTGGTCGTAA TACTATCCCC 3960

TCTATGGAAA TAGTAATAAT AGGGTTTAGT ATTAACCACA TACTTCTTGG CCAACTTGAT 4020

TAAATCAAAA TGGTAATAGG CATCTTCGTA AATCAACCCC TTAGGAAAGG ATAGGGCAGT 4080

TGCAATCTGT CTCTTGATTA GCTTATTGCA AATCGTCCCA GGTATTTTTT CACCTATGAG 4140

GTATTCCTTT AGAAATGTTT GAGAATCACA GACAAAATAG TCATCCTGAT TGGCTGACTG 4200

TGGGCTTTCA TCATTAGCAT AGACATTCAT GACACCACAG CTCGAAACAT CCGCATCTTC 4260

TTGAACTAAT TGCTCATATA AGCTCTGAAT CATTTCTGGA TGGATATAAT CATCTGAGTC 4320

AATAAAAATC AGATAATCCC CGTGAGCCTG CTTCATCCCA TCATTTCGTG CTTGCGACAA 4380

TCCTTCGTTC TTTTTATGAA GCACTGACAC CCTGTCATCT TGTTCAGCGA TTGAATCACA 4440

CAAGCGACCA CTTTCATCTG TTGCACCATC ATCAACAAGA ATAATTTCCA GATTTTGATA 4500

GGTCTGCTTC TGAATGGAAG CTATCGATTT TTCTAGGTAC TGCGCCACAT TATAGACTGG 4560

CACAATCACA CTAATTAATG CAGTTTCCAT GCTACTCCTC TAATAGTTTT TCTACTTGTT 4620

CGATTTGTTT TGTAATTGTA AATTGTTGAA TGAATTGGCT AGCCTCATCG ACATCAAAGT 4680

TTGAGGCAGA AGTCATGTAA TTAGTAATCG CCTGAGCTGC CTCTTGATTG CTCTCAATGA 4740

TTTGTCCAAA TCGTCCTTCT TGGGATAATT CCTCAGCCCC TCCAACGTCC GTAGAGATAA 4800

AAGGGAGTCC CAGACTCAAG GCCTCCACAT ACACTCCAGG AAAACCTTCT TGTTTAGACA 4860

TAGACAAAAG AACTTTCGTC TGAGATAGAT ACTGATAAGG ATTTTTTTGA TAACCAAGGA 4920

AATGTACATA GTCCTCAATC CCATACTCTT TGACTCGTTT TTTCAGTTCC TCTTCCATAT 4980

CACCAGCCCC GATAAAATAG AGATGATAGT TTTTTCCCTC TTGGTGTAAT AATCGTATCA 5040

CTTCCACTAC ACGGTCAGAA CCCTTATTTT CCTCAATCCG TCCGATAGTA CAGATACTTT 5100

GAGGAGCAAT CTCGATATCG ATCTTCTCTT GAGATTTTTC TAGAATAGTC TGAAAATCAT 5160

ATCCATTGTA GATTGTCTGT AATTTAGAAG TATAATCTGG ATAAACTTCC TTGATAGAAT 5220

TGCTGGTCTT TTTTGAAATC CCTACAATTG TATTCGCAGC ATCCAACTGG CTTCTATGTG 5280

ATTCTCTTTT AGAGCTATCC TTAAGAAGTT CTTCAATACT TCCATGAATC CAAGATATCT 5340

TCTTGACTTC TCTTCTTTTA GAGAACAACA GTGGTGGATT CATAATGGTA AAAGAAACTT 5400

CAACATCATA ATCATCTTTT ACAAGCAAAC GACGAGTCAG TCTTGGAAAA TAAATTCTCA 5460

TTCTCCACAA AAAAGCTCGT AACCATCTGG TTTGGCGATA ATCTTGAAGG GATTTTAAAA 5520

TGCGTACATG CTTTGGAACA GATTCATATC CCTTGTCAAA GTGCTCCATT TCAAGAATAT 5580

CAATATCATA CTTTTCTGGA TCCAGATTTG AAACAATGGT TGATAGAATC TTCTCTGCAC 5640

CACCTCCAAG AGAAAAAGAC CACATAAAAA ATAAGATTTT TTTCTTAGCC ACCATATTCT 5700

CCCTTGTATT CTGTATAAGA CTTATCCATA TCAGCGATGA CAGCATCATG ATGCGGTACC 5760

TGCTTGTCTG CTGGTGGAGG CGTCATATAA TCCCCAAAAG CAGTTCTGAG ATAGACATCA 5820

TAGCCGATTG GAATAGGCAT CTCTGTTCCT TCAAATGGCA AGAAAAGATT GTCTTCAAAA 5880

GATGTGATTG GGTACTTGTT TCTCATGTAG CCAGGACCTG AGCATAATTC TGTAATGCCA 5940

TCACAATCAG CCAAATCATA CTTAGTCATT TCTTTCTCAG CTTTTTTCCA GATGCGATAA 6000

CGGAGAGATT TTGGAGTCAA ACCCAGTAAA ATGCGACTTC CCCATTTCAT GAGATCACCA 6060

TGCTTTTCTG GAATAGTTTG CGCACAAAAG AGTGAATAAA TCAAGGCCCA ACGAACCTGT 6120

TTTTTCCGCT CAGCTGGATT TTTCGGATAA TAATCCAAAG GCAAAACATC CAAGGCCAGA 6180

CCATGTGGCA AATCCAAATC CTGCTGATAA GGCTTGATAC AGGTGGTTTT CTTGTCACGA 6240

ATGGTAATAA AAAGATTACG ATCAACAAAA TCCTTGTGAC TCTTTGACAA GAAATAACGT 6300

TCATCTGCAT AACGAGGCCA TAATTCTGCT AATTTCTCAT AATCTTTACG AGGCATAAAA 6360

AAGTCTAGGT CGTCGTCCCA AGGAATAAAT CCCTTGTTTC GAAGGGCACC AATAGCGCCT 6420

CCGCCACAGA GATAACAGAG CAAATCATGT TCTTTACAAA AGGCCACAAA ATATTCAGCC 6480

ATCTCCAGAC TACGAGCCTG AATTGCTTTT AAATCAGTCA TATTGTTCAT TATTCTTTCT 6540

ATCGTATCGT TTCATTATAC CACAAACAAG GGGTGAAAAT CTATTGCAGA CTGTAAAAAA 6600

TCAAAGCCTG ACTGCTATCC AAATAGCTAT CAAACTTTGA TTTTTCTGTC TTATACTCTT 6660

CGAAAATCTC TTCAAACCAC GTCAGCTTCA CCTTGCCGTA GGTATAGGTA ACTGACTTCG 6720

TCAGTCTTAT CTACAACCTC AAAACTGTGT TTTTAGCAGC CTGCGGCTAG CTTCCTAGTT 6780

TGCACTTTGA TTTTCATTGA GTATTATCTT ATCTTAAGCC CATTTGAGCG AGCTTGGTTT 6840

GATATTTGTT TTGATCAACC AGCAGGCCCA AGCCCCCATA AACATCATAG GCATCTACCC 6900

AGTCACCCAG TTCTGGAATC GTCAATTTTT CAATACCATT TTTTGCTCCA TCCAAAACAG 6960

ATAAACCGTT TGTTAGGAGG AAAGTATAGG GTACGTTGGT TGAGGTCATA GCAAAAACCT 7020

TTCCAAGAGC TTCAGAACCA GTGAAAAGTT TAGTGGGATC TTTAATTTGC TCTAAAATTG 7080

CTGTTAAAAC TTGTTGCTGT CTTTTTGTAC GGCCGTAATC TGCCTCATCA TCATCACGGA 7140

AACGAGCATA ATTGAGCAGG GTCGAGCCAT TCATCTGCTG TTTTCCGACT TTAATGGTTT 7200

GGGTTGGAGA CTCAGTCTCG GTAGCGTATA AATCATCTCC GACTGTAGCT TCTGTTAGGG 7260

GACGCCCATT CAATGTTGAA AATTGAGCAT CAATCGTCAC CCCATCAGGG AAAAGCGTGT 7320

CAATCGCTGT GGCAAAGGCC TGGAAATCAA CCAAGGCGTA GTACTTAATG TCCAAGTCAA 7380

AATTATCTTT CAAGACTTGG CGAACCATTT CTGCCCCTTT TTGCCCCTCT TGTTCTCCTA 7440

ACTCGTAGGC TACGTTTAAC TTGTTATCTG TCTGTTTTCT ACCATTAATC ACTTGACTAT 7500

AACCATCTAT ATAGACCAAA TTATCACGCA TGAAACTGAC TAGCTTCATT TTCTTATCTG 7560

AGCCCCCGAC ATTTAATACC ATAATAGAGT CAGTTCGTGT CTCAACACTG TTCTGGCCGA 7620

TTCGACCATC AGTACCCATG ATTAAAATAT TAACTCCATC TCTAGTGTCC TGACCATTAA 7680

AGACTTCTAC TTGAGCTGCC CGGGCATCAG CAGTTTTCTT TGCGCTAGCA TCTTGGTAAC 7740

CACGCAAAAA CATGAATACC ATGGCCAAAG CCACACAGAC CAAAAGTGAA AAAATCACCA 7800

TAAAAATTCG TTTAAGACGG AGCTTCCGTC TTTTCTTTTT TGGAGGGAAA GAGAGTGCTT 7860

GTGATTTGGA TTGTGAGCGA CTCCGGTTCG CATAGCTTGG TAAGTCAACC TGCTCTTCTC 7920

TTTCTTGTTC CAAGCTAGAG CTACTATTTC CCCTAGCAAG AGTTAGCTTT TCTTGCAAAT 7980

AGGCAAACTC ATTTTTTTCT CTCTCATTGA GATAGTGAAT ATTTTTTAGC AAATAATCAT 8040

AACGCAACTG CTCATGATGA CTTAAGGGAT TTTCTTTACT CATCTTCTCT CCTTTCCATG 8100

GTCTGATATT GGATAAATAG GATAGGCACC CAGAATTTTA TACTGGATTC CAATCGCTTC 8160

TAATTCTTTT TGGGCAAAGT GGACCAAGTC CTTATCGGTA TAATCCACAT CGATAATGAA 8220

AAAGTATTCA CCCAGTGCTG TCTTGAGTGG ACAACTTTCA ATTTTTGTCA AGTCAATTCC 8280

TCGCCAAGCA AAGGTCGACA GGGCCTTATA AAGTGCACCT GGAAGGTTGT CAGGTAATGT 8340

CAAGGCCAAA CTCATCTTTT CAGTTTGTGC TTGCAAGGGA ATACTAGGCT TTTCAGCTCC 8400

TAGAACCCAG AAACGTGTGA AATTGGCTTC CATTTCCTGA ATATCCTCGG CAATCAGTTC 8460

CAATCCATAT TCTTCAGCAG AACTTCTAGG TGCAACTGCT GCAAAGGGCT GGTCTGGATG 8520

TTCGGAAATA AAACGGGCCG CATAAGCTGT ACTAGCTGTT ACCTCGATTT GAGCCTCTGG 8580

ATATTGTTCA TCGATGAATT TCTTTCCTTG AGCCAAGGCC TGTGGATGTG AAAAAATCTT 8640

TTCAATCTTA GTATGGCCTG GAACCACCAT CAACTGCTGA TGAATAGGCT GAACGATTTC 8700

TGCTACTGCT TGGATGTGAG CCTGATGAAA AAGATAGTCC AAGGTTTCAT GAACACTACC 8760

CTCAATAGAA TTTTCAACTG GCACCACAGA ATAGTTCACT AATCCTTGCT CATAAGCCTT 8820

GATGACATCT GTAATGTTGG CAAAAGCCTG CAATTCCTCA TGAGGAAAAG CTGTCTGCAC 8880

AACGTGGTGT GAAAATGATC CCTTGGGACC TAGATAAGCA ATTTTCATCT TAGTTCCTCT 8940

ATAATTTCCT CTGGGCTTAG CTTGGTCACA TCCAAAACCC GACTAGCCAC TTCCTCATAC 9000

CAAGCCTGTC TTTCTTGGAA AATAGCTACT AGTTCTTCCT TGCTATTATT TAGAAAAAGC 9060

GGTCGCTGAT TGTCCTTATC AGCTGCGATA CGTTGGTAGA GGGTTTCAAA ATCTGCTCTC 9120

AGGTAGATGT TATCTGTATT AGTCTTGAGT AAGTCACGAT TTCTCTGAGA AATAACCACT 9180

CCTCCTCCAG TTGACACGAC TTGGTCTGTT TGTAGTAAAT CAGCTAGGAC TTCTGATTCT 9240

ACCTGACGAA AGGCTGTTTC TCCCTTTTCA GCGAAAAAAT TCGCAATGGA CATACCTAGG 9300

CGATTCTCAA TCAGAGCATC CATATCAAGG TAATTAGGGT CCAAGCCTCT TGCAATAGTC 9360

GATTTTCCAG CCCCCATAAA CCCTAATAAC ACCTTAGCCA TGAATCAAGC TCTCCAAATC 9420

ATCAAAGAAA CTAGGATAGC TGGTATTGAT GGCTTCTGCA CGGTCAAGCT CCACCTCTCC 9480

ATCTGCAACC AAGAGGGCTG CGATAGCTGT CATCATGCCG ATACGGTGGT CACCAAACGT 9540

ATTGACTCTA GCACCGTGAA GAGCTGATTT TCCTTTGATA ATCATCCCAT CTGCCGTAGG 9600

AGTAATATCT GCTCCCATAC TATTTAAGGC GTCTGCCACA ACCTGAATAC GGTCTGTTTC 9660

CTTGACCTTG AGCTCCTCAG CATCCTTGAT AACTGTTACA CCTTGGGCTT GGGTCGCAAG 9720

CAGGGCAATA ATGGGCAATT CATCAATCAA TCGTGGAATC AAAGCGCCAC CAATCTCTGT 9780

TCCTTTCAAG TCAGAAGACT CAACAATCAA GGTAGCAGAT TTAGCGACTG GATCGATTTC 9840

AGTTATTTCC AATTTTCCAC CCATGGCACG AATGACATCA ATAATACCGG TGCGAGTTTC 9900

GTTGATCCCC ACATTCTGCA GCACTAGACG AGAATTTGGA GCAATCAAAC CTGCGACTAA 9960

CCAAAAGGCT GCACTGGAAA TATCTCCTGG TACGACCACC TTCTGTCCTG TCAATTTTTG 10020

TGGCCCCTGG ACTGTGATTT TCTTACCATC CACACTTAAA TGACCACCAA ATTGTTTCAA 10080

CATATCTTCA GTATGATTAC GGGTGTACTC TTTTTCGATA ATAACTGACT CCCCCTTAGC 10140

TTGTAAGGCT GCAAACATCA AGGCTGACTT GACTTGGGCA GAGGCAATTG GCAACTCATA 10200

ATGAATAGGT CTTAGGTTTT TCGTCCCTTT TAAGCGAAGG GGAGGCAAGT CTCGTTCAGT 10260

TTGCCCTGAA ATGCTGACGC CCATTTTTTT CAGTGGAAGG GTCACACGGT CCATAGGACG 10320

TTTGGAAAGA CTATCATCTC CAAACATCTC TACTTCGAAA TCTGCACCAG CAAGGACACC 10380

TGAAATCAGG CGAATCGAGG TGCCAGAATT TCCCATATTA AGGGCATTTT GTGGCGCTTT 10440

TAAGCCAGCC ATGCCTACAC CTTGAATGGT AATAACCCCA TCTTTATCCT CAATTTCAAC 10500

ACCAAGGTCA CGAAAAACCT GCATGGTCGA AAGAACGTCT TCACCTCGCA GAATATCATA 10560

AACCTTGGTC TCACCCTCAG CCAAACTTCC AAAGATAATG GAACGGTGGC TGATAGACTT 10620

GTCACCTGGG ACGCGGATAC TACCATGTAA ATGGCGAATG TTTGTTTTTA GTTTCATACT 10680

GGACCTCATA CTTGCAATAC TTTTACCTAT TTTATCATAA AAAGCCAGAA ATTCCTTAAA 10740

AATTCCTGAC TTTAGGATCG TTCTTTTCTT ATTTCAGCAA TTCTGAAACT GGTTCAAAAA 10800

CAATTTTTTC AATATCAGAA AGGTAAATGG CCAATTGTTG TTGCTTGGTA AAGAATTCTG 10860

ACAAGAGGCT ATTTCCTTGA ATCTGTTTAC CAAAGCCTTC CATCTTAGCT TGGAAGGACG 10920

CATCTGGCAT TTGACCTGTC TGTGCTAGTT TTTGAATTTC CTCTTGAAAG GCAAGATAAT 10980

CTGTAAAGAT TTTGCTTGCC TCAGCATCTG CTGCAATCGC ATCTTTAGCT GCTTTAACAG 11040

CCTTGTATTC TGGTAATCCG CGTAGACCGC GACTGAGTTC GTTTGCACTA TCGTAAATAT 11100

TTGACATGTT CTTCTCCTTA TTTGATGACG ACTGTATAGT CAGTATTTTC TGTTATGAGA 11160

TGCTCAGCTC TTTCCAAGTC TTGAGCATTT TTAAATGAAA TTTGTAGGAT TCCGTGAATA 11220

TCCTCACGAT TTTCCTCGTT GATGTGGATA TTAACCAAGG AAGTTCCACG TAGCAGTTCC 11280

AAAATCCGCA GGATGACATC TTCTTCATCA GGAACGTCAA CATAGAGGTC GTAAGAGCTA 11340

TCCACACCAC CACGCTTATG GATTTCCATG GTCTGGCGTT GTTCACGCGC TTGGTTAAAA 11400

AAGTTCCAAA TTTGCTCTTC ATCTCCCTTA CTAATGGCCT GACCAATCGC TTCCAAACGT 11460

TCCTTGAAAT CCTCAATTCT ATCCAGAATG ATCTCGCTAT TGGACAAGAG AATGGAGGTC 11520

CACATTCCTG GCTCGCTTTC CGCAATTCGG GTCATATCTC GAAAACCACC TGCCGCAAAG 11580

CGCCTTGCCA TCTCATGCTC TTGAGCATAG ACCGCAGTCT GCTCCATGAG ACTAGAAGCC 11640

AAAATATGAG GAAAATGGCT AATCTGAGAA GTGACACGAT CATGCTCCTT GGCATCAATC 11700

TCGATAAAAC GAGCATGAAG ACCTGAAAGC AGATCCTTCA TTTCCTTAAG CGTGTCCTGA 11760

CTTGTCAGGC TTGAAGGTGT AAAGATATAA TAGGCATTTT CAAAAAGATT GACATCTGCC 11820

GAAGCAGCCC CTGTCTTGTG ACTACCAGCC ATGGGATGGG CCCCGACAAA GCGAACAGAC 11880

TTGCCAGCCA AATACTGCTC CGCCGCATCC ACAATGGTTG ACTTGGTCGA ACCAGCATCT 11940

GAAATAATAA CGCCTTCTCG CAAATCCAAA TTGGCCAACT CCTTAATGAA AGCAATAGTT 12000

TGTTTGATTG GCAAGCTGAG GATAATGACA TCTGCCAAAG GAGCAAAACT AGCAAAATCA 12060

TCCGTTGCAC GGTCAATCAT ACCTTCTTTC AAGGCGATAT CTCTCGAAGC TTGACTACGA 12120

TTATAACCTA AAATTTCATA ATCTGGATGA TCGCGTTTGA TACCAAGTGC CATAGAGGCT 12180

CCAATCAACC CAAGACCTGC GATATAGATT GTTTTTGCCA TAGGAACTCC TTAATAGTTC 12240

TTTGTATAGT CTCGGTGTTT GGCTACCGCT TCTTTTAGTT CCTCAAGATT ATCTGATGAG 12300

AATTTTTCGA GGATTTCTTG CGCCAGAACC GTTGCTACAA CTGCTTCCAT GACCATTCCT 12360

GCAGCTGGAA GAGCAGTCGG ATCACTTCTC TCCACGGTTG CCTTGTAAGG TTCGTGGGTT 12420

TCGATATCCA CACTCATAAG AGGTTTATAA AGAGTAGGAA TGGGTTTCAT GACCCCACGA 12480

ACAACGATGG GTTGCCCATT AGTCATACCA CCTTCAAAAC CACCTAGATT ATTGGTACGG 12540

CGAGTATAAC CGTCTTCTTT AGACCAGAGA ATTTCATCCA TAACTTGGCT GCCTTTACGA 12600

TAACCAGCCT CAAAGCCAAG ACCAAATTCC ACCCCTTTAA AGGCATTGAT AGAGACAACA 12660

GCTTGAGCCA ATCTTGCATC CAATTTTCTA TCCCATTGGA CATAGGAACC AAGACCAACT 12720

GGAACGCCTC CGACGACTGT CTCCACAACC CCACCGATGG TATCACCATC ACGTTTGATT 12780

TGGTCAATAT AGTCCTTGAT TTCCTGTTCT CGTTCTTGGT TGACAATAGA AACTTCAGAC 12840

TGGGCAGCTC TTTGCTTAAT TTCAGCGACT GTCAGATTTT CAGGAACATC GATTTCCTTG 12900

CCACCAAAGA CCACGACATG GTTGGCAATC TCCATATCCA GCTCAGCCAA GAGGCGTTTG 12960

GCTACTGCAC CAACTGCCAC CCGCATGGTG GTTTCACGAG CTGATGAACG CTCCAAAGAA 13020

TTTCGCAAAT CATCAAAACG GTACTTAATC CCCCCAACCA AATCGGCATG ACCTGGGCGA 13080

GGATGAGTAA TTTTCCGCTT GCTTTTAAGG CGGTCTTCAA TGTCCTCCGC AGACATGATG 13140

TCCAGCCATT TCTGGTGGTC CTTATTGATG ACATCCATAG TAATAGGCGC CCCTGTCGTC 13200

TTCCCGTGGC GAACGCCCGA AGTAAAGACA ACCTGGTCAT TCTCAATCTT CATACGACCA 13260

CCACGACCGT AGCCACCCTG ACGGCGTCTA AGGTCCTCAT TGATATCCTC AGCTGTCAAT 13320

GGAAGTCCAG CTGGAATTCC CTCAATAATA GCTGTTAGAC GGGGGCCGTG TGATTCTCCT 13380

GCAGTTAAAT ATCTCATACA CTCTCCTTAT TTTACCAAGT AGTCTTTCAT CTCTTCCAGA 13440

GAAACTGGGT GAATGGTCGC TGAACCAAGC TCTGGCACCA AGACCAATTT CAAGGTGTTA 13500

CCACGCGCTT TCTTGTCATG AGTAAGAGCC TGATAAAGCT TGCCAACTTC CCAATTTTCA 13560

TAGTCAACAG GCAAACCGAA TTTCTGACAC ATCTCTGTGA TAGATTGGGT AATGCCAGCT 13620

GGCATGAGGC CTTTTTCCTC AGCAACCTTG GAAATCTGTA CCATTCCCAT GGCAACAGCC 13680

TCTCCATGCA TGACCTTGCC ATAACCGGCA GTCGCTTCGA TGGCATGGCC AATAGTGTGG 13740

CCAAAATTGA GGTAAAGACG AATACCATTG TCCAACTCAT CTTCAACCAC CATCTTGCGC 13800

TTCACCTGAC AAGAATGTTC AATCAAGGTC TCTGCATGTT CCAAAATACT CTCAACAGAA 13860

CCATTCAGTC CCGTCAAGAG AGCCCACAGT TCTGGATCCT CAATCAAGCC ATACTTGATA 13920

ACTTCACCCA TCCCTTCAAT CAACTCTCTT TTTCCGAGGG TTTCAAGAAC AAGTGGATCA 13980

ATCAGAACCC CATCTGGTTG GGCAAAGGTC CCCACCATAT TTTTAGCAAA TGGTGTATTA 14040

ACGCCTGTCT TTCCACCGAT AGAAGAATCA ACCTGAGCTG TCAAACTAGT CGGAATCTGA 14100

ACAAAGTGAA TACCCCGCAT ATAGGTAGAG GCTACAAATC CAGCCAGGTC CCCAACAACG 14160

CCACCACCAA GAGCAACGAT TCCATCGCTA CGAGTCAGAC CTTGCTTGAC TAGAAATTCA 14220

TAGACTTTCT GAACAGTAGT TAAATTCTTT CTTTCTTCAC CTTCTAAGAA ATCAAAAACA 14280

GCTACCTGAA AACCAGCATC TTCTAGGCTG AGCTTGACCT TCTCTGCATA GAGAGAGGCT 14340

ACATGGTTAT CTGTCACAAT GACTACCTTT TGCGGTTGCC AGAGTTCTCG CAACCACTGA 14400

CCAGCCTGGG CCATACAACC TTTTTCAATC TGAATATCAT AAGGATGGTG AGGAATATCG 14460

ATTCTGATTT TCATAGGAGA GTCTCCCTTT CTTTATTGGT ATTTTTCTGT TAAAGACTGC 14520

CAAATCTCTT CTGTCGGCAT TTCCTTGCCT GTCCACAGTT GAAAAGCTTC TGCAGCTTGA 14580

TAGAGTAACA TTCCCAGACC ATTGACTGCT GGATTGCCCT GACTTCTAGC CCATTTCAAA 14640

AACGGTGTTT CAAAGGGTTG GTATATGATA TCTGCAACTA AAAGAGTTTC TGGTAAGACT 14700

ATGTTTTCAG GAACAGGAGA GGATTGGCCA TCCATGCCCA CACTGGTGGC ATTAACTAGC 14760

AAATCCGACT CGGCAATCCT TGCTTGCAGT TCAGAAACAT ATTCTAAAGC ACACAAATCC 14820

ACTTTAAAAC CTGTCTGCTC CTGTAACTTG TCTAGGTAAG GTCTTGTTTT TTCCATAGAA 14880

ACGGAACGAA CAAAGACCGA AATCTGACTG ACGCCATCCA AAATAGCCTG TGCCAAGATT 14940

GATTTAGCCG CACCACCTGC ACCCAGCAGG GTCATCTTTT TACCTGAAAT TGTAAAAGAA 15000

GGCAAGCACT TAAAAAATCC CTTGCCATCT GTATTATATC CAATTAAATT GCCATTCTCA 15060

TTGACAACCG TATTAACCGC ACCAATCAAG CGCGCTTCAT CGCTCAGCTT ATCCAAATAA 15120

GGAATCACCT GCTCCTTATA GGGCATGGAC AGATTGATGC CAAACATCTG GTAGCGACGA 15180

ATATTGGCCA CTGTTTCTAC CAAGTCACTC GCTTCAATCT CCCAAGCCAC ATAAGCACCG 15240

TTGGTAGCTG TCGCCTCAAA GGCTCTATTG TGGATGAAGG GAGAAATAGA ATGCTTAATA 15300

GGATTGGCAA CAACTGCAGC TAAACGTGTA TAGCCATCAA GCTTCATCCA AAATCTCCCT 15360

GATTTTTTTC ATGCTAGCTA GAGAAATCTG CCCAGGGGCA CTAACCTCAT CCAGACTGGC 15420

AAAAGACCAA CTCGAACCAG TCACATCCGC AGTGATACGA GAGACCTTGC CCACCTTACC 15480

CATAGAAATG GTCACATATT CCTGTTCAGG ATTGAGGGTT TTAAAGCCTC GTGTATAGTT 15540

CATCAAGTCT AAGACATCCT GCTCCGTGTG AGCCATCACC GCAACCTTAA CAAGTTTTGG 15600

ATTTAGGATC GTCAACTCTG ACAAGATTTC CATCATGTTC TCAGGTGTTT CTTGGAAATT 15660

ATGGTAACTC AAAACAAGAT TTGGGAAGTC CAGCATTTCC TCAAAAACAT CCTTGTAGCT 15720

ATAGTACTCA AAATCAATAT AGTCTGGTTG ATAGAGTTGC GCAACTTCCT TGATTAGATG 15780

GATATACTCT TCTGGAGAAA GGTCGATTTC TCCACCTTCG GAGCGAGTTC GTAGCGTGAA 15840

AACCAACTCA CGGCCTGCGA ATTTTTCAAA AATGGCTGGA GCTACCTGCA AAATCGCTTC 15900

TTTAGGCAGA TAGTCGGCAC GCCATTCAAT GATGTCGGCA TCCAGGTACC TCGTGGCATC 15960

CAGAGCCTGA GCCTCCTCTA AACTTCTTGG CATTACTGAA ACGATTAATT TCATTTACTA 16020

ACCTTCATAC TAATCACCTT GAGGTAATTA CTACTTTCAT CTTTTTTATT ATAGGCAAAA 16080

TCTGCTGGAA GACCATATTT GTTTAAAATC TGGTAACTTC TTCCTGCAAA ACCTTTATCA 16140

ATTTGTTCTG TAAATTTCTG ACGGGAAACA TTGGCAGCAT TGGTACTGGC AATGATAATC 16200

CCTCCCGGAT TTAAAATCTC AAGACTCTGG GAAATCAACT TGTGATAATC CTTGGCCACA 16260

GAGAAAGTTT GTTTTTTATT CCGAGCAAAG CTAGGCGGAT CTAGGACAAT CACATCGTAG 16320

GTCAAGTCTT TGCGTTTGGC ATATTTGAAA TACTCAAAGA CATCCATGAC TATAAAACGA 16380

TGCTCGTCTG TGCTGAGCCC ATTTGCCTGA AAATGCGCTT GAGACAATTC TCGTGAACGT 16440

TTGGCTAGAT CAACAGAAGT TGTATGGCTA GCTCCTCCCA TGGCCGCAGC TACTGAAAAA 16500

GCCGCTGTGT AGGAAAACAT ATTGAGTAAG GATTTACCCA TAGCCAAGCC GTCAACTAAA 16560

CTACCGCGAA CCTCATGCTG GTCTAGGAAA ATTCCTGTCA TCAAGCCATC ATTCATAAAG 16620

ACTTGATACA GGACACCATT TTCTAAAACA TTGAAAAAGT CAGGTGCTTC TTGACCATAA 16680

ACATGGGCAG ATTCATAGTC CAAACCCTTA AAGCGGATTT TCTCATAAGC TCCTAAAACC 16740

TCAGGGAAAA CCTGTCTAAA GGCTTCTGAT ATAGTCTGAC GAATCTGATA AACATAAGAG 16800

TTATACCAAG AAAAGACGGC GTAGTCGCCA TAAAGGTCCA CTGTCAGACC CCCAAAGCCA 16860

TCTCCCTCTT GATTAAAGAG ACGAAAGGCA GTTGTCAAAT CATCTTGATA GTAGGCGTTT 16920

CTCTTTTCTT TGGCTTTTCT AAACAACGTT TCAAAGAAAG CTTGATTGAA GGCCACCTTG 16980

TCTTTGCTGA TAAACCAGCC CAAGCCCTTG TTTTGCTGAG AAAGGTAGGC AGTCCCAAGA 17040

AAGTTTCCTT CCTGACCCTG CACCTCTACT TCCTGATCCT TAAGATTGAC ATTCTCAAGA 17100

TCACTGGCTT CTAGTAAAAC TAGCCCCTTA GCAAGCTTCT TTTCAACCCT TTTGCTGACT 17160

CTTATTCTAT TCATAACTAC CATTATATCA AACTTTTAGA CAATTCTCAA AAAAGAAACT 17220

ACCCTTGCTT TTTTACTCTT CTTTTAAAAA ATGGTATACT AGACTTCCTG CAAAACTAGG 17280

AAGTAAATGT GTAAGAATCA CAGTAAAAAA TGCTCTTCCG TCTTGGAGGA GCATTTCTTT 17340

TTATCAACGA AAATCAAATA GCAAACTATG AAACTAGCCT CAGGTTAACT GTGAGATTAT 17400

AGGTAGAGAG GTTGTATCAG CAATATGTGT CTGTCAAATT TAGTGACAAA GGTAGTAGAA 17460

GAAAGATAAA GAAATAAATC AGCTTCAGTA GGTATCTGGA AAATTTGATT TTATAGAGAA 17520

GCCTTTTGTT ACAAACTCAA TATACTATCA ATAAATAATA TTATAGAAGC AACAATAATT 17580

ATAATTTCAC CTATCTGCAT CATTCTATTT CGAACTCTAA ATATATGTTC TATCAAAAAT 17640

ACTTGGAACA CACACATTAT AGGAATTAAC GTTTTTGAAA TTGAAAAATA TCCAAATAAA 17700

TAAACTATAA ACAACAAAAA TAGAACTATG TTATATTTCT TATTCAAAAC ATTCCTCCCT 17760

ATATATTTTT GATTACCAAT CTTAATCATT TACAACTACA TTCTAACAAA CTATAAAAGC 17820

GTTTGTCGAA TTGAATTTAT CAAGCAAGCG ACCAACCAGT TCATCTTTTT TCTATTTCTG 17880

CCAATATGCG TGACAGGTAA TAATGATAGC CAAAAATAGC AAGAGCAAGC AAGACGATAA 17940

GAGCTCCTAC TCCCAAGCTG ATGGCAAGGA TAGGGGAGAG AGACTGAACC AAGAATATGC 18000

TCCCAATTAC AAGGGCCATC AGGATTGCAC TATAAATAAA CAATAAAACT ATGGCGACTA 18060

TGCCATTTGA ACGATTCACC AGGTCCGTAA TGCTACTCCA ATTGGTTGAC AGATTTTTAA 18120

CGTCCTTAAA GTAATGGTGG CAAGAAAGGA TGACACTGGC AATGATCCAG ACTACAAGAA 18180

GGTAAATCAT CGAAATGATG GGCAAGCCTA GATATAGAGA AAGACCAAGC AAAGTCAGAA 18240

CTGGTAAAAA GGACTGGACA GCATATATAA TCCAAAATTT CACTTTCACA TAACGAGCAA 18300

AGTCAAAGGG TAAACTCTTA AGAAAATCAA CATTTTCCCT CTCCAAGGAC AAGGCAATTG 18360

AATGCAGGCT GGTGATATTG TTATTGACAA CTGCTATAAA GAGAGCTATA AAAAACAAGG 18420

GTAACCAGTA TGGAGGATGA ATGTCTGGAA CTATCTGAGA ATCTCGGATT TTGGAAATCA 18480

GACCGATCAT CATGAGATAA GGAAGGAAAG CACTTGTAAA AAGCACTGTA ATCACGCCAG 18540

TCCCCTGTCC CAAGAGGGTG AGGTGGTAGC GTAAAACCAT GCGAAAAAAT CCCTTTTTAG 18600

TGGTTGAAAT TCTCTCCTTG CTGCGACGTT CTTTTTTGAC CTTCTCCTCA CTATTAAGCA 18660

GGATCACGTC ATAAAAACGA GGAAGGACCT TCTTTTTGGT CAGATAAAGC AGGAAGAGAG 18720

TTAGTCCTAT CCAAGCGAGC AGACCCACTA AGGCTTCTGT CGAAAAAGGC TCCACTGCTA 18780

TTTTGTAAAA GATATGAAGA GGATAAAGGA GAAATGGAAT GTCTCTAACT TTGTCAACAA 18840

TACTTCCAAA AGTCGACTGA AGAAAGAAGA TAAATATTAA AGGTATGAGA ACTCCTATCC 18900

CAATCATCAC ATTCGAAAAA ATAGACTGAT ACTTTCTGAA GACCCTAGTT TGAGCCAAGA 18960

AATGCACTGC CACTACCATC ACTAGAGCCA CAGAGACAAA TAATAAGGTC AAGGACAGTA 19020

GCATCAAAGG CAAACCCAGC CATAGAGAAG GAGCTAGCCT AATGTAGAGG ACCAGAAAAT 19080

AAGCTAGGAT TGGTACAATT CCAGTTAGAG CTGGCAAAAG GACAGACAGT CCTTTAGCAA 19140

TTATAATCTC TGATTCTTTA AAGGCATAGG GCCTATACGA TACCAAATCC TTACTCTCAT 19200

AAAAGACATT GTAAAAGGCC GTTAAAGAAG TTGAAAAGGC AATCACTAGT AAAATAGCAA 19260

TCATCGAGCT AAAATAAATA GGTATTTCCT CAAAAGGAAA ATGAATGGCT ATATTACTAA 19320

AACAGATGAT CATCAAGAGA CTGGAAAAAA TGTAAGAACT TAAGACTCTA GCGGAAACAT 19380

TTACTTTTTT 19390

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18436 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

CCGAGCGTCG TTACAGACTT TATCAAGATT GGACGCAAGA AGAAATTCAA CATATAAAGG 60

AAAATATGGC ACAATCTCCA TGGCATACTC ATTACCATGT TGAGCCAAAA ACAGGACTTC 120

TCAACGACCC AAATGGCTTT TCTTACTTTG ATGGCAAGTG GATCCTCTTT TACCAGAATT 180

TTCCTTTTGG TGCAGCCCAC GGTTTAAAAT CTTGGGCACA GCTAGAAAGT GATGATTTGA 240

TTCACTTTAA AGAAACTGGA ATCAAAGTTT TACCAGATAC TCCATTAGAT AGCCACGGTG 300

CCTACTCTGG TTCTGCCATG CAATTTGGCG ATAACTTATT CCTATTTTAT ACAGGAAATG 360

TTCGCGATAA AAACTGGATC CGTCACCCAT ACCAGATCGG TGCTTTGATG GACAAGGAGG 420

GTAAGATTAC AAAGATTGAC AAGATCTTGA TTGACCAGCC AGCAGACTCT ACTGACCACT 480

TCCGCGATCC ACAAATTTTT AACTTTCAGG GTCAATATTA TGCCATTGTC GGCGGACAAG 540

ACTTGGAGAA AAAAGGTTTC GTTCGTCTCT ACAAGGCTGT CAATAACGAC TACACAAACT 600

GGCAAGCAGT TGGCGACCTT GACTTTGCTA ACGACCGTAC TGCCTACATG ATGGAATGTC 660

CTAATTTGGT CTTTGTAGAG GAACAACCTG TCCTTCTCTA CTGTCCACAA GGATTGGATA 720

AGAAAGTTCT AGACTACGAT AATATCTTTC CAAATATGTA TAAGATCGGG GCTTCCTTTG 780

ACCCTAAAAA TGCCAAAATG GTAGATGTGT CTCAACTTCA AAACATGGAT TACGGTTTCG 840

AAGCCTATGC AACTCAAGCC TTCAACGCTC CTGATGGGCG TGCTCTAGCA GTTAGCTGGC 900

TTGGTTTGCC AGATGTTTCT TACCCATCTG ACCGTTTTGA CCACCAAGGA ACCTTCTCTT 960

TGGTCAAGGA ACTCACTATC AAAGACGACA AGCTCTACCA GTATCCAGTC GCTGCTATTA 1020

AGGACCTTCG TGCTTCTGAA GAAGCCTTCT CAAACCGTTC CCAAACCAAG AACACTTACG 1080

AACTTGAACT CAACTTGGAA GCTAATAGCC AGAGCGAGAT TGTCTTACTT GCTGATAAAG 1140

AAGGTAAGGG ACTTTCAATC AACTTTGACC TTGTAAACGG TCAAGTAACA GTGGATCGTA 1200

GCCAGGCTGG AGAACAGTAT GCCCAAGAAT TTGGGACAAC TCGTTCTTGC CCTATCGAGA 1260

ATCAGGCTAC TACTGCTACA ATCTTCATCG ATAACTCTGT CTTTGAAATT TTCATCAATA 1320

AAGGAGAAAA AGTATTTTCT GGTCGTGTCT TCCCACATGC GGACCAAAAT GGTATCCTGA 1380

TTAAATCTGG AAACCCAACT GGAACTTACT ATGAATTAGA TTATGGTCGC AAAACTAACT 1440

GATGTCGCCA AACTTGCAGG CGTCAGTCCT ACTACCGTTT CTCGGGTTAT CAATAAAAAA 1500

GGGTATCTAT CTGAGAAAAC CATCCAAAAA GTCAATGAAG CCATGCGAGA ATTGGGCTAT 1560

AAACCCAACA ACCTGGCTCG TAGTCTGCAA GGAAAATCAG CTAAGTTAAT CGGCTTGATT 1620

TTCCCCAATA TTTCCAATGT TTTCTATGCA GAATTGATTG ATAAATTGGA ACACCAACTC 1680

TTCAAAAATG GTTACAAGAC CATCATCTGC AACAGTGAAC ATGATTCTGA GAAGGAACGC 1740

GAATACATCG AAATGTTGGA AGCCAATCAG GTGGACGGCA TCATTTCTGG TAGTCACAAC 1800

CTAGGAATCG AAGACTACAA TCGTGTGACA GCGCCGATTA TTTCCTTTGA CCGAAACCTA 1860

TCGCCAGACA TCCCTGTCGT CTCCTCTGAC AACTATGCTG GTGGGGTTCT TGCTGCCCAA 1920

ACCTTGGTCA AGACAGGTGC CCAGTCTATC ATCATGATTA CAGGGAATGA CAATTCTAAT 1980

TCGCCAACCG GACTGCGCCA CGCTGGTTTT GCATCCGTAC TCCCAAAAGC TCCTATTATC 2040

AATGTTTCCA GTGACTTTTC TCCCGTCAGA AAAGAAATGG AAATCAAGAA TATCTTGACC 2100

CGGGAAAAAC CAGATGCCAT TTTTGCTTCG GATGATTTGA CAGCTATTCT GGTCATTAAA 2160

ATCGCTCAAG AATTGGGCAT TTCTGTCCCA AAAGAGCTCA AGGTCATCGG CTATGATGGG 2220

ACCTACTTTA TCGAAAATTA CTACCCTCAA TTGGCTACTA TCAAGCAACC TTTGGAAGAG 2280

ATTGCTTGTC TCACTATTGA TCTTCTCTTG CAAAAGATTG AAGGCAAGGA AGTCGCCACA 2340

ACTGGTTACT TCTTACCAGT TACGCTATTA CCAGGAAAAA GTATTTAAAC ACAAGAAAAC 2400

TCAGACCGAT TCGTCTGAGT TTTTATGATC TTAAATTTTC GAGATAGCGC TGGGCTGTCT 2460

CTAGGTTAAA GGTTTTATCT GAGATGAGGC GCTCTACTAG GGGAGCAACT TCAGATTCAC 2520

TAGCCCCAGC TAGGAGAGCT AGGGATTTGG CCTGTAGTTT CATGTGGCCT TGCTGGATGC 2580

CCGTACTTAC CAAGGCTTTG AGGGCTGCAA AATTTTGAGC AAGACCGATG GACACGATAA 2640

TCTGGGCTAA TTCTCTGGCA GAAGGATTTC CTAGTAGATC ATGACTGAGA ACTACACGTG 2700

GGTTGAGGCC GATAGAGCCA CCCTTAGTCG CTACAGGCAT GGGCAGGGTC ATCTCACCGA 2760

CCAATTCTTC TCTTTCAAGG TCCAGCGTCC AGCAGCTAAG ACCTTGATAG CGTCCATCTC 2820

GACTGGCAAA GGCATGGGCC CCAGCTTCGA TGGCACGCCA GTCATTACCA GTGGCAATCA 2880

AAATCGCATC AATACCATTA AAAATTCCTT TATTATGAGT AGCAGCTCGG TAAGGATCAG 2940

CCTGCGCAAA CTGACTAGCC AACGCAATTT TCTCCGCAAT CTCTCGTCCT TGATCCTTTT 3000

GGCGGCTCAA GTAGCGAAAG GCGATGCGAC AGCTTGCAGT CACCAGAGAA TCGGTCGCGT 3060

AGTTGGACAG GATTCCCATG AGACTCTGTC CCTGACTGAG TTCTTCTAAG ACTGGTTTCA 3120

AGGCTTCCAG CATGGTGTTG AGCATATTGG CACCCATGGC TTCCTGGGTA TCGACATGAA 3180

TATAAACAAC GAGAAAGTCT GGTTCGCCTT TTATCTGCTC GACATGCAGA TCACGCGCCC 3240

CACCTCCACG TTTAACGATA GAAGGATAGG CTTGATTGGC AAGCTCCAAG AGCTCCGCTT 3300

TCTTGCTGGC AATCTTCTCT TGCGCTAGTT TAGGATTAGC AACTTGATAA AGGGCTACCT 3360

GCCCAATCAT CTGTCGCTGA TGGACTTGTG CAGTAAAACC ACCTGCACGC TTGATGATTT 3420

TGCTGGCATA GCTGGCCGCC GCAACCACAG AGGGTTCTTC TGTCACATAG GGAACGGTGT 3480

ATTCCTGACC GTTGACAAGT ACCTCCGGAA CCAGTGAATA AGGCAGAGAA AAAGTTCCCA 3540

CTACATTCTC ACTCAGCTGG TCTGCCACAG TCACGCTCAT CTGTTCATCC TTCTCCAGAC 3600

TAGCTTGTCT CTCAGGACTA AGGAGCGCCT GAGCTTTTAA CAGCTCGAGG CGCTCTTGGT 3660

ATGATTTTTT AGAAAATCCA TTCCAACTTA TCTTCATTAT TTTTCAACCT TGCTATAACG 3720

GCGTTGGTGG TCGAGAATTT CAACCAAGGC AAAATCTTGA TTTTCATAGC CAGCAAACTG 3780

GGCAGAGTTA GTTTCATCCA AGTTTACTTC CTCAAAAAAG ACCTTTTCAT AGTCTGCAAC 3840

GGATAGGGCA GTTCGTTGGT TGAGCTTGTT CAAACGGTCT TTATCCAAAT AAGCTTCATA 3900

TCCTTCAACC AATTCACCAC TGAAGAACTC AGCCACAGCT CCACTTCCGT AACTATAAAG 3960

GGCGATTTTA TCCCCAGCTT TCAAGCTATC TGTATTTTCC AAGAGAGACA AAAGTCCAAG 4020

GAAAAGTGAA CCTGTGTAGA TATTCCCCAC CTTTTGACTG TAGAGAATAG ACTGGTCAAA 4080

ATGCTTTTGT AAGAGGTCTT TTTTCTCTTG AGGCAGGCTC TTATCCATGA TTTTTTTCAA 4140

GCCTTTTAGC GCTAATTTAG GATAAGGCAA GTGGAAACAA ACAGCCGCAA AATCATCCAA 4200

AGTAAGCTGG TAGCGTTTTT GATATTCAAG CCAAGTCGTT TTCAAACTAT CCAAGTATTG 4260

TTGGGTAGAA TAGACACCAT TTACATAAGG AGTTGTCGAG TAATTTGGTC GCCAGAAATC 4320

CATGATGTCA CGGGTCTGAG CTACATTGTC ATTATTAAAG GCCATCATGC GTGGATTTTG 4380

TGTAATCAAC ATAGCTACAC TTCCAGCACC TTGAGTTGGT TCTCCTGGAG TTTCAATACC 4440

GTATTTGGCA ATATCACTGG CAATGACCAA GACCTTGGAC TCCGGAGAAT TTTCCACATG 4500

CAATTTGGCA TAATGGAGGG CAGCAGTCGC TCCGTAGCAG GCTTCTTTAA TCTCGAAACT 4560

ACGAGCAAAG GGCTGGATGC CCAGCAAGCC ATGCACAAAG ACGGCCGCAG CCTTACTCTG 4620

GTCAATTCCT GACTCGGTCG CCACAATGAC CATGTCAACT TCTTGTCTTT CTTGCTCAGT 4680

TAAAATAGAG TCACTAGCAC TGGCCGCCAA GGTCACGATA TCCTCAGTTA GGGGCGCAAT 4740

ACTCAATTCC TTGAGTAAGA GTCCTTTACT TAATTTTTCA GGGTCAATTC CCCTCGCTTC 4800

TGCTAAGTCT TGTAATTTCA AGACATATTG ACTGGTCGCA AAACCAATCT TATCAATACC 4860

GATTGTCATA TTTACCTCTG TTTTATCATT CATGTAAAAA ATCGTTCTAT ACTATTTTAT 4920

CACAAATGGC AGTAAAAGAG AGAAAAAAGA CTTGATTCAC CAAATCAAGC CTCTTATTGG 4980

TCATCATTTT AAAGAATGAT TAGTTGCTAG AGAGTTCACC GATATAAGTA GCTTTATAAG 5040

CTCCATTCAC AGTTATCAGC TCCTGGAGGA TCAAATTTCC TGAGTAAGTC CTTCCCATCT 5100

CATCTACAAA TTTTTGATAA AACTGACTGG TCGGAATTTC TCTGACATCC TTATCAAATG 5160

TCTTATCAAG TGTTTTACTA ACCTTCTCAG CAATCAATTG ATGCTCTTGC CATCCACTTT 5220

GAAACTCTGA GCCCGAACTA GAAACCATGA CTGGGATAAA CAACAAGGTC AGTAGATTTA 5280

CAGACAATAA GGAAAGTAGT AGACTTCCTG CAAAACTAGA ATCCTAGTTC ATGATTGATA 5340

ATACCAGCAA TCAAATTCAT TCGTAATCCG AAGCGTTTAC GATGATTTCG ATAGGTTGTT 5400

GAAAACATTT TAAACGTTTT TACTTTGGCA AAGATGTTCT CAACCTTGCT TCTCTCCTTA 5460

GATAGCGCAT GGTTACAGGC TTTATCTTCA GCTGTTAGCG GCTTGAGTTT GCTGGATTTA 5520

CGTGGAGTTT GTGCTTGAGG ATATATCTTC ATGAGCCCTT GATAATCACT GTCAGCCAAG 5580

ATTTTACCAG CTTGTCCGAT ATTTCTGCAA CTCATTTTGA ACAACTTCAT ATCATGACTA 5640

TAGTTCACAG CGATATCCAA AGAAACAATT CTCCCTTGAC TTGTGACAAT CGCTTGAGCC 5700

TTCATAGCGT GAAATTTCTT TTTACCAGAA TCATTCGCTA ATTGTTTTTT AGGGCGATTG 5760

ATTTTTACTT CCGTCACATC AATCATTATC GTGTCCTCAA AGCTGAGAGG AGTTCTTGAA 5820

ATCGTAACAC CACTTTGAAC AAGAGTTACT TCAACCCATT GGCTCCGACG GATTAAGTTG 5880

CTTTCGTGGA TACCAAAATC AGCCGCAATT TCTTCATAAG TGCGGTATTC TCGCACATAT 5940

AGAAAGCGTT ATCAATTTAT TTATCTCATT TTTCAGAAAA TTCTTTTATT TCTGTAAAGT 6000

CTACGATACT CGATGTGTTT TTATATAATG ATAGAGTCTG AGAATCACTG TTCCGCTAGC 6060

CATTCCAATA GAGATTACCA AAGCCAACAT GACAACCAAG GTCGCACTTG CCAGTGCTTT 6120

ATTATAGTCC CCTGTCACAA AAAAGGCAGT TGTTCGGTAG GAGAGATAAC CTGGAACCAG 6180

CGGTGCCAAA ATGGCCAAGA TAAAGACCAC AGCAGGTGTC TTATAAAGAA TACTTAAAAT 6240

CTGGCTGACA CAAGAACCAA TAATGGCTGC AATGAAGGTA GCTACAATGA CATTGGTCGG 6300

TTCCTTGAGC AAGAGATAGA TTAGCCAGAC AGTCATGCCC AAAATCCCTC CAGGTAAGAG 6360

CATAGACCGT TGCACATTGA GTACGATTAA AAAAGTGATA ATGGCAAGAA AACTTGCTAC 6420

TGCTTGTAAT AAAAAGGTTG TTAGTGTCAT ATTAGTTCAT CAATACCAAG GCGACAGAAG 6480

TTCCTGCCCC TAAAGCGAGG GTAATGAGCA GGGATTCAAA CATCTTACTC ATACCAGAGT 6540

TTATGTGGTT GGTCATAATA TCACGGACCG CATTGGTCAA GGCAATACCT GGTACAAACG 6600

GCATGACCGC ACCAGCTATA ATCAAATCTG CCGTTGAAGG AAAACCTGTG TAGCGAGCCC 6660

AAAACTGGGC AATTATCCCA AAGACAAAGG CTCCAGCAAA GGCTGTCACA AAGGGAATTC 6720

GGATAAATTT TTCCACATAG AGGGAAAAGG CAAAACCAAA TAAGGTCGCC ACTCCTGCCC 6780

CAAGTGCGTC GTAGATATTT CCGCTAAACA TAACTGAAAA GAAAGGAGCA CTAAAGGTCG 6840

CAGCCAGAGT TACCTGCAAC TTAGTATAGG GAAGGGGTTT GGCTTGCAAG GCCGTCAATT 6900

GCTTAAAGGC TGTTTCTAAG TCAATCTGCC CCCCAACTAG CTGACGAGAA ATCTGGTTCA 6960

CATCGCAGAC TTTTTCGATG TTATAAGAAG AGGAGGTCAC GCGCTTCATG CGCAAATATT 7020

GGTATTTTCA ATAGAGAAAA AGATAGCGGC AGGCATGGCA AGGACATTGC AATCCACAAT 7080

CCCCTGCGAA TGCGCGATTC GAATCATGGT ATCTTCTACA CGATGGATTT CTGAGCCACT 7140

TTTAAGGAGA ATAGTCCCCG CTAGCATAAT CACATCAATG ACGGCATTTA ATTCTTTTGA 7200

TTCTTCCATG CTTTCCTCCT TTTATCAACT CCCTCTATTC TATCACAAAT CCGGACTCAA 7260

AAAAAATCTT TGCCATGAAA TCATGACAAA GATTGATTAC TCATTTTGAT TATCCATCTG 7320

CTTTTAAGGA GTAGCTGAAG TTGTTTTAGG TTTGTAGATT GAAATCTTGA CTCTAGTCTT 7380

ATTGAGGTCT ACCTTTTCAC CTGCTCTAGG ACTTTGTTCA ACAACCATGC CTTCTGCACT 7440

ACCTGCAGGC GCTGTCGTCA CTTCTACAAC TTCTATATTA GCTTCCTTAA TCCCAACAAT 7500

TTGAATCAAA TTGTTCTTAG TAAACTCCAA GCTAGAACCA ATGTAACTCG GCATGGCAAC 7560

ACTTGTAACT TTTTTAGCTA CTGTCAAGAC AATTTGAGTA GGTTTACTCA CATCATAAGT 7620

CGTTCCGGCA CCTGGACTTT GTTTCATAAT CGTTCCTGGT TCGCTTTCGC TGGACTCTTC 7680

TTCCTCTATC TTAATCAAAT TCTCAGGAAC CTTCTTCTGC TTGAGTTCTG AGATTACTTC 7740

TGTAGAGTTC CGTCCAATAT AGTTCCCTAA TTGAATCGTC GTAGCTTTTT TAGCTACTGT 7800

CAAAACAATT TGAGTTGCCT TGCTCAAGTC ATAGGTCGTA CCTTCTGGTA GACTTTGCTT 7860

CAGGACCGTT CCAGCCTCAC TCTCATTCGA CTCTTCTTCC TCAATTTTAA TCAAATTATC 7920

TGGAACTTTT TTCTCTTTTA ATTCCGCAAT GACATCAGAG GATTTCCGAC CGACATAATT 7980

ACTAATTTGG AAAGATTGCT TGCCTGATGA GACAACCAAA TTGATTTTCG TTCCTTCTTT 8040

TCGACCAGTT CCAGCGCCAG GATCTGTACG GATAATCCGC CCTTCTTCCA CCTTTTCACT 8100

AGCCTCTGTC TTCTCCTCAC CAATCTCAAA ATTGGCTTTT TTGAGCGTTG CCTTGGCCTC 8160

TGCAACTGTC TGACCTGCCA CATCTGGAAT GGCAATGGTT GCAGGAGTTC TGGATAGTAT 8220

CCAAATAAGA GAAGCTGCCA CCAATACAAG GCTGGCCAAC AAAATCAGGT AACGCATCTT 8280

AAATCTATGT TTTTTCGGTG CTTGTGGTTG GTAAGTTTCC TCTGTCACAG CCTGGCTTGG 8340

GTTTTTGATT GATTTGTGTT CTGTTTGCGC TTGAACCTTA GGAATAGATG TCAAGGTACT 8400

CTGAGAAACC TTCGGCAAGG TCTTGGTATC TGCCTTGCTC GTTTCATCAA AGATTAACTT 8460

ACTTTCATTT CTACGATTGT AGGACAAGCT ACTAGACAAG TCCACATACA TCTCTGAAAC 8520

CGAGCGGTAG CGATTGGTCA ACTTTTTAGC AGTTGCCTTG ATAATAACAT TTTCTAAAGC 8580

CTGAGGTACA GATGGATTTT CTGCAATAAC GGACGGCAGG GGTTTCTGGA AATGCTGGAG 8640

GGCAATGGTC ACCGCGCTAT CCCCGTCATA AGGGATATGG CCTGTCAGCA TCTCATAGAA 8700

AATAATCCCC ATGGCATAGA TATCACTCTG CACAGTCGCC TTCGAACCAC GCGCCTGCTC 8760

TGGTGACAAG TAATGAACTG AGCCCAACAT CGAGTTAGTC TGGGTCAGAC TTGTCTCTGC 8820

AAAGGCTACA GCAATCCCAA AGTCTGTGAC CTTGGCAGTC CCATCTGGTG TCAAGAGGAT 8880

ATTTTGAGGT TTCAAGTCCC TGTGAACAAT TCCTCGAGTA TGGGCCAAGC GCATAGCCAA 8940

GAGAATTTGT CCCATGATAC GGACTGCTTC TTCATTAGAA AGAGGATAAT GTTCCTTGAT 9000

ATAGCGTTTG AGGTCCAGTC CAGCCACATA CTCCATAGCT AGGTACTGTT GACCGTCTTC 9060

CTCGCCAATA TCTGTTATCC GAACGATATG AGGATGGTCT AGATCTGCCA TAGCTCTCGC 9120

TTCACGCTGA AAACGAGCTA CAGCTATCGG GTCCGTCTGG TAGTTGGTCC TCAGAACCTT 9180

CACTGCCACT TCTTCCCCAT CTAAGATTAA GTCTTTGGCT AGGTAGACAT CCGCCATACC 9240

TCCTCGACCA ATCTGTTTGA CAATCCGATA GCGTCCGGCA AAAATCTTGC CGATTTGGAT 9300

CATTCTGCAT CCTCCTCGTT CATAGAAACA AGGGCAACCG TAATGTTGTC TAAACCTCCT 9360

GCATTGTTAG CAAAACGAAC AAGTGTCTCC GTTTTATCTG CTAAAGGAAT ATCACTGGTT 9420

ACAATATCAC GAATCTCACT GCCTGAAATC ATGTTGGTCA AGCCGTCACT ATTGAGCAAG 9480

AGATAGTCAC CTGACTCAAG GATAACTGTC CCAAAATCAG GCTGAATTTC ATCTTTTTGC 9540

CCAATAGACT GGGTGATAAT ATTTTTTTGC GGATGAGCTT CTGCCTCTTC TGGTGTCAAT 9600

TGACCAGCCT TGAGCAATTC ATTAACCAAG GAATGATCGC TCGTCAACTG ATGGTATTCT 9660

TCTCCACGAA TCAAGCCGAT ACGCGAATCA CCAATATGAG CATAGATAGC CTGATTATCA 9720

ATAATAGCAA GGACTTCCAA AGTAGTTCCC ATGCCTCTGT AAGCTTCATC CTGACCAAGC 9780

TGGTGAATCT TTTGATTTTC AATTTCTAGG TAATGGGCGA ACCATTCACG CACTTCATTG 9840

ACTGTATCGA TCTGGGTATC AACCCAAGCT ACACCCAGGT CTGTGACCGC CATTTCACTA 9900

GCGATATTCC CTGCGCGATG ACCTCCCATC CCATCAGCTA AAATAATCAT GGTACGTCCA 9960

GCTCTATTGA CATAGTGGTT GACATAGTCT TGGTTATTTG TTCGTTTCTG ACCAACATCT 10020

GTTAATAATG AAATTTCCAT GTGTCAGTTC CTTCCTAATC CGATATCTTG CGAAATTGAC 10080

TGATGAAGAA TCCATCACTT CCATACAATT CAGGTGTAAT GAGGATACAG CCGTCTTTCA 10140

TGATATCCTT ACATTCATGT TCTAGTTTTA CCTGCTCGAA CTCGGGATGA CTCTCTAAAA 10200

AGGCCTTAAC GACTTGAAAA TTCTCCTCTG AGACGATAGT GCAGGTGCTA TAAGTTATTA 10260

TACCACCTTT GCCTAGTATT TGACAAACAC TACCTAATAT TTCTAACTGA ATTTCCTGCA 10320

AGGACGCGAA ATCTGCCGTT TCTTTATTGT ATTTGATATC TGGTTTTCGG CGCAAAAGAC 10380

CGATTCCTGA ACAAGGAGCA TCCACCAAAA TCTTATCAAA GGAATCCTGG TCAAAAAACT 10440

CATGCACCTT TCTGGCATCC AATTTTTGAG TTTGAACCCG ATCTTCAACT CCCAGACGTT 10500

GGGCATTTTC TTGAATTAAA TCCAACTTGT GGTCGTACAA GTCCAGAGCA GTAACCTGAC 10560

CTGTCGTAAG ATAAGAGGCT ATATGGGCTG TTTTCCCACC TGGAGCCGCA CAGGCATCAA 10620

GCACTCGCTC ATCACCTTGT AAATCAAGCG TCGGAGCAAC CAGCTGACTG GACTCATCTT 10680

GGATGGTAAT GGCTCCATCC GCAAACAAAT TATGCCCTGC AAAATGCCCC TGCTCCTTAA 10740

CCAGACCAGT GGTTGCTAAA AGGGAATTAT TCGCCTCCAA CAAGGCTTGG ATTTCCTCTT 10800

TTCGACTTAG GTCTGTTACA CGAATACTGG CTTTGTTTCG CACTAACAGG CTTTCAAAGA 10860

TGGCTTTTGC TCTCTCCTCT CCGTATTCTT CCTTGAGTTT GGCAACTAGC CAAACTGGGA 10920

GAGAATAGGC AATGGAGTCA CGCTTGTTTT TTCGCTTGAT GCTAGCAATA TCTGGCCAGC 10980

CTTCACGCAA GATACGGCGA AGGACAGCGT TGACCAATTT TTCACTGCCT TTTTTACGGA 11040

700

GTTTGGCCAA TTCCACTGCT TCATTAACCA CAGCATGATC TGGAATCTTG TCCAAATAGC 11100

GGAGTTGGTA GGCACTCATG AGAAGAAGGA CATAGAGCCA GCTGTCTAAC TGGTCTCTGT 11160

CTTCGATAAA GTGGGATAGG TACCATTCCA GAGTCAGTTT ACGGGCTACC GTTCCATAGA 11220

CCAGCTCGGT CACTAAGCCC TTGTCTGCTG CCAAAAGTTG ACTTCCCTTT AGATGCTTAT 11280

TTAAGGCGAT ATTTGAATAT GCTTGGTTCA CAAAAACATC CTCTAGCACT GCTAGAGCTA 11340

AACTTCTAGC CGTTTCTACT TTAGTCACCA AATCGTTCTC CTACAGTCAA TGTACGTCCA 11400

ACTCCGTTGA GGAAGGAAGC AATGTCCATC TTAGGCTTAC CAGCTGGCTG CACTTGTTTG 11460

AGGGATAGAG CCCCTTCAGC CGTTGCGACA ATCAATTCTT TCTTGCCGAT AGAGAGAATC 11520

TCACCTGGAT TTCCCTGACC TTCTACTGGT AGGGCTTCAT AAATCTTAAA GCGGTCGCCC 11580

TTAAGGAAAG TATGGGCAAC AGGCCAGGGG TTCATTCCAC GAATTTGGTT AAAGAGTTGA 11640

CGATTGGTTT TGTTCCAGTC CAGTTTTTCT TCCTCTGGCT TTATATTTGG AGAGAAGGTA 11700

ACCTGACTCG TATCCTGCGG TTCAGGTTTG ATATCACCAG CAATATAGGC AGGCAGAGTG 11760

TCCAAAAGCA AATCACGACC AACTAGCGCC AATTTTTCAA ACAAGGTGCC AACATTGTCC 11820

TCATCTGTGA TCGGAATGCT GCGACGAGAA ATCATATCTC CTGCATCCAT TTCCTTAACC 11880

ATTTCCATGA TGGTCACACC AGCTTCCTCA TCCCCTTGAA TCAAGGCATA ATGGATAGGC 11940

GCACCACCAC GGTGTCTAGG AAGGAGGGAG GCATGAACGT TGACAGCAAA GTCCATGCTA 12000

TCAAGGAGTT TGCTTGGGAG AAACTGCCCA AAAGCAGCAG TCACAATTCC ATCTGCTCCT 12060

AGCTTCATAA GATCTTCCAT CTCTGGACTT CCAGATAATT TTTCAGGTTG GTAGATAGAT 12120

AGTCCTGCTT CCTTGGCAGC CTGCTTGACT GGGGTTTCTT GGATAACTTT TTTACGACCA 12180

ACAGCACGGT CTGGCTGGGT CACAACGGCT AGAATTTCGT AACGGTCATC TGTCAAAAGT 12240

CCTTTTAAGA CTGTTGCTGA AAAGTCGGGG GTCCCCATAA AGATTAGTTT TGTCATATCT 12300

TCTCCTTCTT ATAAAAATTG CTGCGGCTCA TGGTCAATGC TGAGACGGAG CTCACTATTT 12360

TCCCGTTCTT GAGTCAAGGC TAAAACCTGG TTGAGGGTCG ACCCCAGCTC ATCTTCTAAA 12420

CGGTATTTAA TTAAAATCTG GTAATGATAG AGGTTGTGGG TACGGGCAAT CGGTTTTGGC 12480

GTTGGCCCCA GAATGGGACT GGTCTCTGAC AAGCCTGACC GCAAAATGTT CATGACTTCA 12540

TAGGCACGTT TGAAAACCTC TTCTTCTTTC TTGTGAGAAA GGGTAATACC AATCGTGAAA 12600

TAGTAAGGTG GATAGCCGAG TTGTCGTCTG ATTCCCATTT CATAGGCATA AAAGCCTTCG 12660

TAATCTTGAT CCTTGGCAAA TCGAATAGCA TAGTGCTGCG GATTGTAGGA CTGTATCAAG 12720

ACTTGACCTG CCTTTTCAGC CCGACCTGAT CGACCTGCCA CCTGAGTCAA GAGCTGGAAG 12780

GTTCTCTCAG AAGAACGGAA ATCAGGCAGA TTCAAGGCCG TATCCGCATT TAGAACTCCG 12840

ACTAGGGTAA CATTGGGAAA ATCCAAACCC TTTGCAATCA TCTGAGTACC AAGTAAAATA 12900

TCCGCTTCCC CTCGCCCAAA CTGGTCAAGC AAGGCTTGGT GACTGCCTTT CTTTCGAGTC 12960

GTATCCACAT CCATCCTCAG AATGCGAGCT TGGGGAAAGA GTTCTGCTAG CTCATCATAA 13020

GCCTTCTGAG TTCCCGTCCC ATAGTAACGA ATACTGCGGC TCTTACAGTT AGGACAGACC 13080

TGAGGAATAT CCTTCGAGAA ACCACAATAA TGGCAGTTCA TAGTCTTGGT ATCCATATGC 13140

AAGGTCAGAG AAATATCGCA GTTGGGACAA GTATCCACCG TCCCACACTC CCGACACATG 13200

ACAAAGCTAG AATAACCACG GCCATTGAGC ATGAGAACCA CCTGCTCTTT TTTAACCAGA 13260

CGGTCTTGGA TAGCCTCTAG CAAAGGAGGC GTAAAGTTTG ACGTCTCATT TTGTCCGATA 13320

TAGTCTCGAA AGTCAATCAC TTGAACCTCA GGGATTGTAG CCAAAGGATT GGCACGTTGG 13380

GTTAGACGTA AGTGTTGATA GACGCCTTTG CCAGCACGTG CCCGGCTCTC TAAGCTCGGC 13440

GTTGCAGATC CAAGTACCAG AGTTGCTTGA TTATACTGAG CCCGTAAAAT AGCTACCTCT 13500

CTGGCATGGT AACGGGGATT GCTGTCCTGC TTATAAGCCG CTTCATGCTC TTCATCAATA 13560

ATCATGACAC CCAGATTTTT CAGAGGAGCA AAGATAGCAG ATCTGGCACC AACAACAACT 13620

TGGGCATCGC CACGCTCCAC CTTGCGCCAT TCATCATACT TTTCACCATT GGATAATCCT 13680

GAGTGAAGAA TGGCTACCTT GTCCCCAAAA CGTGCTATAA AACGCTCGGT CATCTGAGGA 13740

GTCAAGGAAA TCTCAGGTAC CAGCAAAATA GCTGTCTTGC CCTTATCCAG GGCACCTTGG 13800

ATAATCTGCA AGTAAACCTC GGTCTTCCCA CTTCCTGTAA TCCCTTGAAG TAGAAAGGGA 13860

GGTTGAGAAC TGCCAATAGA ACTCACAACC GCATCACGCG CCTGTCTTTG TTCTGGATTT 13920

AACTCCAAAG GTCTACTTGC TTCAATTCCT TCAAAATAAG CAGCCGAGCG TTGAACTTCC 13980

TTTTGGACTA TGGTAACAGC ACCTTGATCC ACAAAGAAGT TGACTTGCTC TCGCGAGTAG 14040

GACTCTAACA AGCTAGCCAA GGAAGCGCTC TCTGGATGAG ACAGCAGATA ATCTCTCAGT 14100

TCCAACTTTT TCTTGGCACG TGTAGAAATC TCAACACCTT CTAATTGAGC ATGGTCAACC 14160

TCATACCAAG ACTGGGTCTT GACCTTCTTT TGATCGACTG CCTGATATTC CAGACCAAGC 14220

AGGCCTTTTC TAGTCAAACG CATCATTTCA GCTTGCTTGG CAAGGTCTAG TGAAGAAAAG 14280

GCTAGCGAAT CTTCTGAACC AAACAGGCGC ACTCGTTCTT CCTGACTCAA GCCTTCCAGA 14340

GGATAGAGAA TCTTGTCATA GCTAGAATTC AGAAACCCTG GAAGCATGGC CTTGAGGATA 14400

GAGATTTTGT AGGAGAAGAC AGATTTGCGT AACTCCTCAG CCAGCCAGAG TTGTTCTGGC 14460

GTGAGAACAG GAGAAAAATC CAGCACCTCT GCAATATCTT TTAAATCTTG CTCCATCTCT 14520

TCTCCATCTG ATTGGGACTT CAAACCAAGA ACAATCCCTT GAATCAGGCG ATTACCCTTA 14580

CCAAAAGGCA CATGAACCCG CATCCCAACT TCCAGCATTC CCTCAAATTC CTCCGGAATC 14640

CTGTAACTAT AGGGCTGGTC CGTCTGCATC AAGGGCACAT CTACGATAAT CTTAGCTAGG 14700

GCCATCTTCT CACCTCCTCC TTGTCAGTAC ATTCTTGCAA TAGAAAAAAT AAGATTGAGT 14760

CCCCCCAACC TTAAATTTTT TCACCATCTT CTTTTTCTTT AGCAATTTGC TCTTTGATTT 14820

TCTTTTCTTC TTCTTCTTTG CGGCGTTTTT CTTCTTCGAT ACGGCGACGC ACTGCTTCAC 14880

GTTTTCCTTC TGGATCTGGG TGAATTGTAA CGTTTCCTGA TTCGATTTCT TCTAAAGCGC 14940

GAAGAGTTGA TTTTTCAGAC TTGAAACCTT GAGTTGCTGG GGCACCTGCT TCCAATTCGT 15000

GGGCACGTTT TGCTTCCAAG ATTACGAGTG AATATTTTGA AGGAACCTTG TCGAGCAAGG 15060

TATCAATAGA GGGTTTTAAC ATCATTTGCT TGTACCTATT TTCTAAATTT TATCGGGTAG 15120

TTGGAGATTT TGGTAACATC TCCTGATAGT GACCAATGAC ACGATCCACA CAGAAGTGTT 15180

CTGCTTCAAT CACACATTTG ACACGTTCAG CAGCTAGGGG TACCTGATCG TTGACAATCG 15240

CATAATCATA CTCACGCATG AGGGCAATTT CTTCCTTGGC CTTTTCGATT CGTTGGGCAA 15300

TCACTTCTGC ACTATCTGTT CCACGACCTA CCAAGCGATC TTGCAATTCA TCCAAATCTG 15360

GTGGTGTCAG GAAGATAAAG ACAGCATCTG GAACCTTTTT CTTGACCTGA AGAGCACCCT 15420

GAACTTCAAT TTCAAGGAAA ACATCGATTC CCTTGTCCAA GGTTTCATTG ACATAGGTCA 15480

GAGGAGTTCC ATAGTAGTTA CCGACATATT CTGCGTATTC CAACATCTGT CCTTGACGAA 15540

TCAGCTCTTC AAATTCTTCA CGAGTACGGA AGAAATAGTC AACACCGTCC ACTTCTCCAG 15600

GACGTTGTGC GCGTGTCGTC ATCGATACAG AATATTGAAA TTGGTTTTCA GAACTCTCAA 15660

AAATCTCTCT TCTAACCGTT CCTTTTCCAA CCCCTGAAGG ACCAGAAAAA ACGATTAGTA 15720

AGCCTCGGTC TGCCATTGTG TCTCCTTTTA GTCAATCTGT GAAATAACAT TTCTCTAGAA 15780

TAATGGCAAA AAGCCAGATT ATCCTTTACA GTCTTTCTAT CTAGTGTAAC AAAAAAGCAG 15840

TAATTTTTCA ACTGCTCTTT CTTATTTATT TAGCATAATC TACTGCACGA AGCTCGCGAA 15900

TCACGGTTAC CTTGATATTT CCTGGATAAT CGAGATTGTT TTCAATTTTC TTACGAACTT 15960

TGTGAGCCAA GATTGTGACT TTGTCGTCCT TGATTTTTCC TGGATTGACC ATGATACGAA 16020

TTTCACGTCC TGCTTGAAGG GCAAAGCTAG TTTGCACTCC TTCAAAGCCG TTAGCAATTT 16080

CTTCCAAATC ATGGAGACGC TTGATGTAGC TTTCAAGAGA CTCACTACGA GCACCTGGAC 16140

GGGCTGCGCT CAAGGCATCT GCTGCAGCGA CGATAACTGC TATCACGCTC TCAGCTTCAA 16200

CATCTCCGTG GTGACTAGCA ATCGTATTCA CCACAACTGG GGGTTCCTTG TACTTACGGG 16260

CCAATTCCAT ACCGATTTCA ACGTGGCTAC CTTCAACCTC ATGGTCAATG GCTTTCCCGA 16320

TATCGTGAAG GAATCCAGCA CGACGGGCAA GAGCCGCATT TTCACCAAGT TCGCTCGCCA 16380

TGATACCAGC CAACTTAGCA ACCTCAATCG AATGGCGCAA AACATTTTGT CCATATGAAG 16440

TACGGAACTG CAAACGTCCC ATAATCTTCA TCAAGTCTGG ATGAAGGTTT GGCGCACCAA 16500

TTTCATAGGC AGCAGCCTCA CCGTATTCAC GAATCTTATT GTCAATCTCT TGACGGTTTT 16560

TCTCAACCAA CTCTTCGATA CGAGCTGGAT GTATACGACC ATCTTTGAGC AACATTTCCA 16620

TAGTCATACG GGCAATCTCA CGACGAATCG GATCAAATCC TGACAAGGTC ACCACTTCTG 16680

GTGTATCGTC GATAATCACA TCGACCCCTG TCAAACTTTC AAAGGTACGA ATGTTACGAC 16740

CTTCACGACC AATAATGCGT CCCTTCATAG TATCGTCTGG CAGATGAACT GTTGAGTTTG 16800

TTGACTCCGC TACATATTCA CCAGCGATAC GTTGCATAGC TTGAACCAAG ATGTCCTTGG 16860

CCATTTTGTC AGAACGTTCC TTGACCTCTT GCTCAGCTTC GCGAATGCGA CTGGCAATCT 16920

CCCTGGTCAA GTTTTCCTCT GTCTGAGCCA AGATAATATC TCGTGCTTCT GCCTGAGACA 16980

GCGCACCAAT ACGCTCTAGT TCTGCTTCTT TTTGTCTTTC GACTTCCTCT AATTGCTCTT 17040

CACGCGCATC AAGGTTTTTC GCTCTATCAG AAATACTTTG TTCTTTTTGT TCAAGTGTTT 17100

GTTCTTTACT CGTCAAATTG TCGTCCTTAC GGTCAAGGCT AGTAGCTCTC TCTGTCAAAC 17160

GACTTTCGAT TTGTTTGAGT TCTTGACGTT CTGATTTGAA TTCAGCGTCC ACTTCTTCAC 17220

GGTATTTTCT GGCTTCTTCT TTGGCCTCCA ATAGTGCTTC TTTTTTAAGA GACTTGCTTT 17280

CACGTTTGGC TTCATTAACA AGTAAATCCG CTTCACGCTC AGCTTGTCCA CGTAAATTAG 17340

TTGCTTCTTG TTCAGCATTT AAAAGCATCA ACTCTGCAGC TTCCTGAGAT GATTTCATCT 17400

TAGCTGAGAT GCTGACATAT CCAATGACTA AACCAATGAT GACGGCAAAA ACAGCAATCG 17460

CAAGCGACAT GATTTCCATG TTTTTACCTC ATTTTATTGT TATTCCGAAT GACATACATT 17520

CTTTTACATT CTACCATAAA AAAGTGATTT TCACAAACCT AAAATAGAAT ATGTTTTGAG 17580

GAATTTGGAA CACATTTACC AAAATAAACT TGTTGTTTAG AAATAGTAGT TTAGTAGAGA 17640

CTTGAGAAAA AGCCTACCTT TCAATAGACT TAGTAATGAT CTTTAAAGGA CAAGAAAGCC 17700

ACGCTATCTC CATCCATCAT ATAAATCAAG CGATTTTCTG CATCAATACG CCGTGACCAG 17760

GCTCCTTGGT AATCATATTT GAGTGGTTCT GGTTTACCTA TTCCTGTAAA GGGATCACGT 17820

TGAATATCCT TGATTAGTTT ATTGATTCTT TTTAACGTTT TCTTATCCTG ATTTTGCCAG 17880

TAGCAATAAT CTGCCCAGGC ATCTTCTGTA AACTTGAGCA GCATTTCTTA CTCCTCAATA 17940

ACATGGACCT GAGTACTTCC AGCACGAACT TGAGCCATTC CTCGCAAAAC CTTATCAGAA 18000

AGTTCCTTAT TTTGAGCAAT TCTCAGGGTT TCTTGGATAC TATCCCACTC ACTCTTTGAA 18060

AGGACTACAA TGTCCTCATC TGGATTTTTA TTGACCACCG TCAAAGGCTC AAATTCATCA 18120

TTTACCTTCT TCATGTAGTC CTTTAAATGA TTTCGGAATG TTGAGTAAAG GACTGCTTCC 18180

ATAACCATAC CTCGTTTTAG CTCTTTTCCA CTATTATACA CGAAAAGAAA GAAATTGTCA 18240

GGAACTTGTA CAAGATTTTC TTTTCTATCT ATTTATACTC AATGAAAATC AAAGAGCAAA 18300

CTAGGAAACT AGCCGCAGGC TGTACTTGAG TACGGCAAGG CGACGTTGAC GCGATTTGAA 18360

TTTGATTTTC GAAGAGTATT ATTCGTAAAA AATCTCAAAA AGCCTACCTT TCGGTAGACT 18420

TAGTTTGTTT CTATTC 18436

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7001 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

ACGTAGAAAA ACTATTTCTA TCACAGATAA TATTCCGTAT GTTGTTGGAG GTATTGAAAT 60

AAACGTCCTA GGTATCTTTC TCAGTCTATG TGACTTACAA GGGAAAACTC TTTTCGAGAC 120

AGAAATTTTG AATGAAGATT ATCCTATTTC AGAAATCAAT TCCACCATTA CCAATATGAT 180

AAAAACAGCT ATAGAGTACG TCCCTTTGGA AACAAAATTA CTTGGATTTG GCTTATCAAT 240

ACCTGGACAT TATAACAAAG ACTCCGGAAG TATCATTACA AACAACCCCA TATGGGAATC 300

TTTTAATTTA TTAAATGTAA TTAAAAGATT CAATTTTCCT TTTATTGTAA AAAATAATAT 360

CGATTGTATG GCTATAGGAC AATACCTTTT TAATCCACAC AATACCCCCG ATAACTTTAT 420

TTTCCTACAC GCTGGATTAG GTATTTACAC TTCCTTTTTC ACAAAAGAAA AAATAGGAGC 480

CTCTAAAAAT CCTTATATCG GAGAAATTGG ACACACCATT GTCGAATTGA ATGGGCAATA 540

TTGTGAATGC GGAAAAAAAG GTTGTTTACA AACATATATT TCGGATGCTT GGTTAATCAA 600

ACACGCCCAA TTATTATTTA AAAATTCCCA ACTAACTGTA CTAAAAAGCC TTGTAAAGAC 660

TGAAAAAGAC ATTCATTTAG ACACCCTTTT AACGGCTTAT AATTTAGGCG ACTCCGCTTT 720

ACGTCAACAA ATTGATAAAG GAGTCAATTT ATTAGCCACT TCTATTGCAA ATCTCCTCCT 780

CATCAATCCT GCTGATAAAA TCTATATCAA CAGTCAATTG CTTAATTATC AACCTTTCAC 840

TCATGAAGTC AGGGATAAAA TCCAAGACCA GCTCCACTTC GTTCCCTTTA CTCGTAATAT 900

AGAAATTGAA ATTTTACCTT ACAACAAACA TCGTGGAAGT ATAGGAGCTT GTGCATTAGC 960

TATCGTCGCT TTTTTCATAG AACATAGCAA TGTATTACAA GATATTATTT CACCTTAATA 1020

TATTAGAAAT CTATAGACCT GTTTAAATCA ACTATAACCT GTAGTAGATA TCTCGTATTT 1080

AGACAATATG AAAACAAGAC GACTTCCATA TAGGAAACCG CCTTCTCGCT ATGTTGAGTG 1140

ATTTATATTA AAATAACTTT TCTTCTAGCT GCATTTTATT ATTATAAAAA CATTCATCAT 1200

AACCCCCAGA ACTTAAATAA CAATTTTTAT TCAAGATACA TACTCCTAGA ATAAACTTTA 1260

TATGAAATTC TCATTTTTGT TTTTACAATT CTCCTTAGTT AAATCTTGTT TAATATATGT 1320

TTTACATATA GTATTTAGCG CCACATAGTA CTGAACTCTC TCCAAAAAAG GTTATTCCTC 1380

TTTGAATAGG GCGTTATCAC AAGAAAAGCA TCTCCACGTT TCAACTTCAT ATGGCTCAAA 1440

AACAATCAAT TGATGCTAAA ACCTGTACCT AGATGTTTCG GTTCATAAAA CCATGAAACT 1500

GTAAAAGTGG ATGAAATTGA TAGCGATAGT CAAATCAAGA GGCATCATAA CTCTAAAAAG 1560

TCACAATATA TAAGTTCATC CTCGGAAAAA TATCATTCTA ATTGTTGAAA TGCCTACATG 1620

AAAAGAAACG TCAAATGCTC ATGAAACAAC GAATACAGGT ATCAAAACTA TGACAAAACA 1680

AATCCCTAAA TTTACTAAAG ACACTGCTCA ACTTTACACC TGTAAATGGT TGTTGTATAA 1740

TAAAGTTACA AAGATGTACG ACCACACTGT TGTAAATCAT AGTGTTCGCG AATATATTAC 1800

TGATAGCATT TCTACAAATA CAAGTAAAGA GAGCGGATGA GATTCAAACG AAATATGTCA 1860

GTGCTTTGGC ATTCCTAGCC TTCATATCAT TTAAAGAATT CTATAGACAA AATTTTTTCC 1920

AATACAGACA CTCGTAACAA CTGCTTCATT TTTCTACCAA CATATTTAGG AACAGGATAA 1980

GATACAAGAG TATTAATCCA TAGCTCAGTT CTATACCAAT CTAAGACAAA TAAGCTAAAA 2040

AAACGATTGA TAATAAGCAA ATAGATTCCA AATTTTCTCT ATCTGCTCAT TTTAATAAAC 2100

AATACTAGTG TAACTATCCT TCCAGTCAGA AGCTTGTCAA ATCACACCGA AAATTCTTCT 2160

AAAATTTATC TCGTTAGGCA ATCAAGCAAA AACTCGACGA TAGTACAAAC ATTATCATAC 2220

AGGATTGACT TCCTAAATTA TATACTTTAG TAAGGTTTTC GGATAAGAAA AAAGGTTCAT 2280

TTTACATTTC TAAACATTCT TTTCTAAGAT GAAAAACAGA ATTTTTCGAT TGTGATTTAA 2340

AGCAACAAGA AGATTTTCAG TATCATCCTA TAGATACGAG CTAATTAAGA AAAACTACAT 2400

TTTTGAATAT AAACTACAAT AATATAAACT AAATTTTATA GGAGGAAGAC AATGGATTGG 2460

TACGATTATA TGATACAGGC ATCCAAACAA TCACAATTCA ACGCAAGCCA TTGGTTTCGC 2520

TATTTGCGAA AAGTTATTTT TGAAGACTAT TCTTATTTAA CAAACCAAGA TGTAGAAAAG 2580

TTGCTAGACT CCAAAGAACT AACCCGTTTT CAAAAAATTA GCTTGAAGTA TGCCTTTCAA 2640

GAGCATACTC CAACTCATAA ATATGTGATT TCATTAAATA AACCTGCTAA GTTAACCAAT 2700

GTTCAAAAAT TGATGGAGAA ATACAAACAT GGATAAAATG AAACCGGTCT TCCAAGCCCT 2760

AAATAAGGAA TTAATTCAGG AAAATCTGAC TTTAACAATT ATCTGTGTCG GTGGTTATGT 2820

CTTAGAATAT CATGGTTTAC GTGCCACACA AGATGTTGAT GCTTTTATGG CTCTATAATA 2880

TTTGTAGTGG GTAAATCCCC TATGGATATT ATGGAGCCTA TTTTTGTGTA GAAAAAAAGT 2940

CCCATATGAC CTATAATGAA AAGCGACAAA ACAACTCATT AGAAAGAATC ATATGGAACA 3000

ATTACATTTT ATCACAAAAT TACTAGACAT TAAAGACCCT AATATCCAGA TTTTAGACAT 3060

CATCAATAAG GATACACACA AGGAAATCAT CGCCAAACTG GACTACGACG CCCCATCTTG 3120

CCCTGAGTGC GGAAACCAAT TGAAGAAATA TGACTTTCAA AAACCGTCTA AGATCCCTTA 3180

CCTCGAAACA ACTGGTATGC CTTCTAGAAT TCTCCTTAGA AAACGCCGTT TCAAGTGCTA 3240

TCACTGTTCA AAAATGATGG TCGCTGAAAC TTCTATCGTC AAGAAGAATC ATCAAATTCC 3300

TCGTATTATC AACCAAAAAA TTGCGCAAAA GTTGATTGAG AAGATTTCTA TGACCGATAT 3360

TGCTCATCAG CTGGCCATTT CAACTTCAAC TGTCATTCGC AAGCTCAATG ATTCTCACTT 3420

TGAGCATGAT TTTTCGCGTC TTCCTGAGAT TATGTCCTGG GACGTTGAAA CAGTCCGGGG 3480

AGTGACTGTT TCAATCGGGA GATGGAGATG AGCTTTATTG CGCAAGATTT TGAAAAGCTC 3540

GATATCATCA CTGTTCTTGA AGGTAGAACA CAAGCTGTCA TCCGAGATCA CTTTCTTAAA 3600

TATGATAGAG CCGTCCGATG TCGCGTCAAA ATTATTACTA TGGATATGTT TAGTCCTTAC 3660

TATGACTTAG CTAGACAACT TTTCCCGTGT GCTAAAATCG TTCTTGATCG CTTTCACATT 3720

GTACAACATC TTAGCCGTGC TATGAGTCGT GTGCGTGTCC AAATCATGAA TCAGTTTCAT 3780

CGAAAATCCC ATGAATACAA GGCTATCAAG CGCTACTGGA AACTCATTCA ACAGGATAGC 3840

CGTAAACTCA GCGATAAACA TTTTTATCGC CCTACTTTTC GTATGCATTT AACCAATAAA 3900

GAGATTTTAG ACAAGCTTTT GAGCTATTCA CAAGACTTGA AACATCACTA TCAGCTCTAT 3960

CAACTCTTGC TGTTTCACTT TCAGAATAAG GAACCGGAGA AATTTTTCGA ACTTATCGAG 4020

GACAATCTTA AGCAGGTTCA TCCTATTTTT CAGACTGTCT TTAAAACCTT CCTCAAAGAT 4080

AAAGAAAAGG TTATCAACGC CCTTCAACTA CACTATTCTA ATGCCAAACT GGAAGCGACC 4140

AATAATCTCA TCAAACTTAT CAAGCGCAAT GCCTTTGGTT TTCGAAACTT TGAAAACTTC 4200

AAAAAACGGA TTTTTATCGC TCTGAATATC AAAAAAGAAA GGACAAAATT TGTCCTTTCT 4260

CGAGCTTAGC TTTTTTTCAA CCCACTACAG TTGACAAAGA GCCGGAAAAA GGAACAGCCT 4320

TAGCTTTCCT TTCATTTCTT TTTATTTCCC TCGTAGTAAA CGTGCTAGCT TCCACAAAAC 4380

AAACAGGATT CCCAGAAATG CCAGTACCAC TAGCCCACGG TACAACCATT GAGAGGTTGC 4440

AACACGCGAT ACAGATTGTC CTTCTTTCGT AAAAGCAACC CTCGCAACTG CAGCTGTTTG 4500

TGGATCTGAT TTTTGATAAA CAGCGACTCG TTCAAAATTC ACTAATAAGC GTTTATTAAA 4560

GGTAGGAATC GGATCGCAGG TTATCAAGGT CATGATATTT TTAGAGCTAA CCGATTCTAA 4620

TTTTTCCCAT TCCGACGGTA AAATAATCTC TGTGTCCATC ATCTGATATT CTACAATTTC 4680

CTGGCCATTA TCATAATAAA GAGCATCTCC AACTTTTAGC TGATCCAAAT GGCGGAAAAA 4740

GACATGGCTT GGCTCTGCAC GGTGCCCAGC AATCACTGAG CGAATCCCTG TACCATCCAG 4800

AGGCAGCGGT GTACCATCCA CATGAGCCAA GCCCATCCCT AAATGATGAT AATCTGCTCC 4860

CAAATAAACC GGCTCCATGA TTTCCAAACT TGGAATAGAC AAGTAACCAT AGACTGCATC 4920

AGGGTCGTCA GACACTTGGT AATTGACCTC ATATCCCTCC GCCAAAAAAG GATCTACAAT 4980

GCGATTTTGC GAAGCCAAGC GTTGATTGTA GGCGAGAGAA TGGTTCTGTT GTTCTTGGTA 5040

CATTTCAGTT GTCATGGATT TCACAAATGT AGCATGACCT TTCACCTGTC CAAGAGACTG 5100

CAACACCATC TGTCCAAAAC AATAAATAGG AATCAAACAG GCTACCAACA TCAACAAGTA 5160

TCCCAATAAG GCTCGTAGTT TAGTCCTTGA CATGACGCCC CTCCAATTGC TTTTCTAGTC 5220

CTTTGACAAT CCGTCGATTA CGATACACGC GATACAGCAA GAGAAGGATG ACCGCCATCG 5280

CTCCTAGTAA TAACCACAAC CAGAATTGCC CACGCTCTCT CACCGCTCGA TTCCGCTCTG 5340

CAATTGGTGC CGTATACGGA ATCCGCTTCC CACGTACCAA CAGACGATGA CTGTTAATCA 5400

TATACGGTGT ACAAGTCAAC AAGGTCGCAT AATCTTCCCC ATGTTGAATC AAGACAGGCT 5460

CAAAGTCATT CGGCTCCACC GTCACTATCT GATCCACTTG GTAGGCCAAC ACCTGATCTA 5520

AAACGTGAAG ATAAAAGATA TCCCCTTTTT TCATCTTATC CAATTGACTG AACAATTCTG 5580

CCGTTGGCAA TCCTCTGTGA GCAGTGATCA CTGTATGGGT ATTTTCACCT CCAACAGGCA 5640

GCGAAGCCCC TTCTAACAGC CCTGCCCCTT TCTGAAGAAT GTCCTCACTC GTTCCGACAT 5700

ACATCGGAAT TTCCTGATCA ATCGCAGGAA TTTCCACATA GCCAATCCGC TCATGGACCT 5760

TTAGCATATT GGCATATTCT GAGACGCCTT TCTTTTTCTC TTGCTCTGTA AAAGGATCAA 5820

GAATTTCAGA TGGTTTCAAG GTCGCATTGA AGGCTTGAGC CAAGCGCCAA CGCTCCTCAA 5880

GTTCTGCCTT ATCCATCTGG GAAACCGTCT CATCAAACTC TTTAATAACC TCGTTTGACT 5940

CAATACGATA ATAATAACGA GACACCAATG GATATATCGC AACGGCGAAT CCTACTAAGA 6000

AAATCAGAAG AAGGATCAGC GGATGTTTCT TCTTTTTTGT GCCTTTTTTT CGTGAACGTC 6060

TACTGTTGTC CATCCTCCAC CTTCACTTCC TTCCTTGCTG CTTTCAGCGC CTTCAAAGCC 6120

TTTTCCGGTT GTTTTTTCTT CTTGCGCAAG CGTCGAATAA TCCATAAAAG AATCACAATC 6180

AAACCAACTG CCACATAAAA CAGGTAGCGA TAGAGATGAC TGAGTTTGTT TGCTGCAATA 6240

AATTCTTCCT CAACCTCTGC TACGTACGGT ATCCGATGCC CCCGAACCAA TAGACGATGG 6300

GTATTGATCA TGTATGGCGT ACAAGTCAGC AAGGTCACAT AATCATGACC TGGTACAATC 6360

AATAAATCAT CAAAGTTCGT CGGCTCAATC ACCTTTACTT GATCCACTTG ATAGGCCATC 6420

ACTTCCTTGA TATTGTGCAC ATAAAACTTA TCCCCAACTT TAAGTTTGGT CAAATCCGTA 6480

AACATCTTAG CTGTTGGCAA ACCTGTATGT GCCGTAATCA CCGCATGGGT CGAATTGCCT 6540

CCGATCGGCA GAGAAGTTCC CTCTAGATGC CCAGCCCCTT GCTGCAATAC CTCTTCAGCA 6600

GTACCAGCAT AAACCGGCAA ATCCACGTCA ATAACGGGGA TTTCCACATG CCCCATCCGC 6660

TCATGGATTT CTAACATACG TGCATACTCT GCTCGCCCTT TTTTCTTCAT TTCTTCCGAC 6720

CAAGGATCGC CACTCACTAC ATTATTCAAA GAGTCATTGA AGGCTTGTGC CAATTTCATT 6780

CGTTCATCAA TGTCAGCCTC ATCCAACGTT GCTTTTTCCT TATCAAAGTC AGCAATTTGT 6840

TGATTTGATT CCACTCGATA ATACAAGCGA GACACCAGCG GATACGCCAT TACCGCCATT 6900

CCAATGAAAA ATACCACTCC TAATAGGAGA TTATTTCGTT TTTGCTTTTT TGTTTTTACC 6960

ATTTTTATCA GCATCCCTTT ATCTTCAAAC TTCAGGGTAT C 7001

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10411 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

GAGGGAGCTT AAGAAGTTAC CACCGTCCTC TAGCGCCTTA TCCGCATCAA AGTTAAGGTT 60

GATATTTTTA AAACTGTCGC CAGCTTGTGA TACGATGCTT TGTTTAAGGT CATTTAGGGT 120

TTTAGTGAAA TCTGCATTGC TGAGGATATC ACTCTTTGAG AGATTCAAGG CAAAATTGAT 180

GATGATATTG ATCTGGTTTC CTGTTATGAC CTGATCAAGT TTGTAATTTT TTAAGGTATC 240

TTCAACAATC TTGCGGATAT CTTCTTCTGT CAGATTTCCC TTACTTTCTT TAGCTTTGGC 300

GAGTCCTGAC TTGATATCAG CTAGGGCAAC GTTTAATTTA TTAGCATCAT AGCCTGATTT 360

GTCCTTGTTT TCAGCATTGA TATCTGACAA AGCTTTTAGC TCTTCTTGAG CCAAATCTTT 420

ATTAGCTTGT GGCACCTTGG CTCCATTAGC CTCTAGCGAA TAGTAAATCC CTGCTAAAGC 480

ACTTTCTCCT GTAACTGGAA TAGGGGCTGC TACAGTGATT TTGGCATGTT CCATACCCAG 540

CGTTACTGCT GCGTTTCGGT ACATATCCTG AGTCACCTTA GTGATATTTT CTGGTGTTTC 600

AATCTTGACC TCAAGTGGCG ATTTGTCACC TAGCTTTTGA ATCTTGGCTG ATGAATACAA 660

CTGTAAGCTA GAGTCATTGG CCACATTCAT GATTTTAGAA TAAACATCAG GTGTCATGGT 720

CTTGAGTTCT TTGGTATCTG TTGAGGCATT GTAGCCCAGT TTTTTAAGAG TTTGATTTTT 780

TTGGTCTTCA GATAGGGAGG AACCTAGGAC ATATTCAGGT TGGACATAGG TTTCATCGAT 840

AACTTTTTGA ACATCTGTTG CTGCATGGAC GCTATTCATA GCTGTTACTG CCCACAAGAT 900

CGCAGCGCTA GTCAGAAAGA GTTTCTTTCT CATAGGGAAT TTCCTCCTTT ACTTCTTTAG 960

AGTAATATAT CTATCTTAAA GAAAACTTAT AACAAAAACA CCTGGTCTAG CCAGATGTTG 1020

AAAAGAGAGT GAAACATTTG ATGATGTAAA GGTTAAGTCG TACCTGTCTA GAATAATAAT 1080

AGTTTCCTCC ATTTACATAG AGTTCAGCAC CGTGAAAAAT GGAAATGGGG TGAATATAAC 1140

TATAAGTCTT TCCAGTCCTA TTACCAAGCA AGGGGGCAAC AGTCTCACGA GAGTACTGTT 1200

TGGCTAGAGC CAGGGTATTT TCCTTGCCAT TTTGGGCGAT AAAATCGATA TAGGCAGGTC 1260

CAAAATTATA GGCTTGAACA GCTGTCCAGA TATCTACCCC CTTCTTCTGC GCCAGATAGA 1320

GATTGCCTGT CAGAGTTTGA ATGCCTTGCC GAATGCTAGA GGCATTATCA TTGATGGTGT 1380

TGGTGGAACC ACTTGCAGAC TCACTAGACT GCATAACATC GCCTTCTTTT CCTTTTGTTT 1440

CAGTATAAAT CATAGCAAGC ACAAGCTCTT CGTTTGCTGG GGTGTCTTGT TCACTCAATA 1500

TTTCTCGCAC CATGGGTTGA TAGGTCATGA CTTGTTTGAC ATCTTGATGA ACGCGGTAAG 1560

CTTTATAGCC AGCAAAAAGG AAGACTGCTA GTACAAGCAC TCTTCGAATT CGTTTAAACA 1620

TTATTTACTT TGGATATCCT CGATATTTTT GATTAAGATA GAGTAGGTTC CATTTTCGTT 1680

TTGGATAAAC TCAACAGACT CGGCGTCTTG ATAGACGTTA TTGGGAACGA TGAGCTCAAT 1740

TCCATTTGAT AAGGAGAGTT TTTGGTTTTC AAATTTCTTT AATTGGCGAC TGGCATCAAT 1800

TTCATCAAAT TGAACAGGTT CTGGTACGGC TTCTTTGACT TGGTCAATAA AGCTCAAACG 1860

AGCCGTCAGA TTGTTGTCAA AAAGGTCATT AGCCAATTTC TCAGGTGACA ATTCATTGCT 1920

TTCTTCTAGG TTGTTGAAAA TAGCTGATTT GACCTTGGAT TGAAATTGAA AATCATCTGT 1980

GTTAAAAGAT TTAGCAATTC TCTGGGCTGT TTTTTCCAGT TCCTTGATAG ATTTTTTAGG 2040

AGAAATCTTA GGAGCGACAG CAAGAAGATT ATCTGAAAAA TAGTTCAAAA AAGTCCCGTT 2100

GTACTTGATT CGTTTTTCAA TCAGGTGATA CTTGCTACTC TGAAGATTGA CCACCAAGGC 2160

CTCATCAGCT CCTGTTCCAA ATCCAGGCAG GTTATTCTGA GTTAGCTTGA TTGGATTATC 2220

AACTTCTCCT CCGAGGTGGG TCAAGGTCTC CCGCAGGGCA ATTCGCAAGA AAGCGAAATG 2280

TTCTACACCT TCTTTAGAAA ATTGCACAAA AATCAAGTCA TTGGTCTTGA GATTTTCAGA 2340

AATGCTAAAC TCCTCTTTCC AGAGATTAGC CAGCGTTACT GATGTCTCCA ACAAATCGTC 2400

TGTAATATGA TTGAAGAAGG GATTTTCTTC TTCGAAAATC CCAGTCTTGG CTTCATCTGA 2460

ATACACATGT TCAATTTTTT TACGCAGGTA TTCTTCGATT TTTGGAGTAA TATTGAGAAA 2520

CTTATCTGCT AAGAACAGTT CGGTATCATC CGGACTGAAC TGGTGAATAA TGGCTTTCTT 2580

AATATAAATG TCCATAAAAG TTTTAGTCCT CGTATAATGG GAAGGCATCT GTCAATTCTT 2640

TGACTGCACT TCTCACTTCT TCTAATACAG CCTCATTTTC TGAATTCTTA AGGGTTTTAA 2700

TGATGAGTTC AGCCACTTTG CGACTTTCTT CTTCACCAAA TCCACGTGCA GTAATGGCTG 2760

CTGCTCCGAT ACGAATCCCA CTTGTCTTGA ATGGTGACAA GCTTTCGTAA GGGATTGAGT 2820

TTTTATTTAA GGTAATATTG ACTTCATCCA ACAAGTTTTG AGCAACTTTG CCGTTTTCTA 2880

CAACTTTAGT CACATCAACA AGGAAGAGAT GGTTTTCAGT TCCACCTGAA ATAATACGGA 2940

AATCAGGGTC TTGCAAGAAG ACATCTGCCA TAGCCTTGCT GTTCTTAATT ACATTGGCAG 3000

CATATTCCTT GAAGGCTGGA TCCAAAACTT CTTTGAAGGA AACTGCCTTA GCCGCCACAA 3060

CATGCTCTAA AGGACCGCCC TGAATACCTG GGAAAATAGC TGAATTGATT TTTTTAGCAA 3120

GTTCTTCGTC ATTGGTCAAA ATCAAACCAC CACGAGGTCC ACGAAGGGTT TTGTGGGTCG 3180

TTGTTGTTGT GATATGAGCG TATGGAACTG GGCTTGGATG AAGGCCAGCC GCAACCAAGC 3240

CAGCGATATG GGCCATGTCC ACCATGAGCT TCGCACCGAC AGCATCTGCG ATTTCACGGA 3300

ATTTTGAAAA ATCGATAATT TGAGAATAGG CTGAAGCACC AGCTACAATC AGTTTTGGTT 3360

TTACTTCTTG GGCTTGTTTC AAGATAGCAT CAAAGTCTAA GAGTTCCGTT TTAGGATCAA 3420

CACTATAAGA AACAAAGTTG TAGGTTTGAC CAGAGAAGCT AACAGGAGCC CCATGAGTCA 3480

AATGACCACC TGATGCCAAA TCCATTCCCA TAACCGTATC ACCTGGCTCA ATCAAGGACA 3540

TGTAAGCCGC ACAGTTAGCT TGGCTTCCTG AATGTGGTTG AACATTGGCA AATTTAGCAC 3600

CGAAAATTTC TTTTGCGCGT TCAATAGCAA GAGTCTCTAC AACGTCTACT ACATCAGTTC 3660

CACCATAATA ACGGCGTCCT GGGTAACCCT CGGCATATTT ATTTGTCAAG ATAGACCCTT 3720

GAGCTGCCAT AACAGCCTTG GAAACTACGT TTTCCGAAGC AATTAACTCG ATATTATTTT 3780

GTTGGCGTTC TTCTTCTTTG GCAATAGCAT TCCAGAGATC AGCATCATAT GCTTTAAAAT 3840

CATCTTTGTC AAAAATCATA GGTCTTCTCC TTTATTGTGT GACTAGTCCA TTAGTTTGAT 3900

TTTACAATAA GAAAATCAAA CTAACAGATG CGAATAAACC GTTTCTGCAT TTTATCACAA 3960

GTATAGCCAA CTTTTTCATA AAATGCATGA GCACCCAGAC GATGATTGGC AGAATTTAAG 4020

CGGATAAACC CATAACCACA TCTTTTTGCT TCTTCTTCCA ACCCTTGTAG TAAACTTTTA 4080

CCAATACCTT GACCTTGCGC TTGAGGTGAA ACTGCTAAAG CTAAGATATT AAATCCTGCT 4140

TTGGAATAGA GTGATTCGTA AACTTCAGCG TGGACATATC CAAGTAAGAC ATGATTAGCT 4200

GCATCCTCAT AGCCAAGTAG GAAATGATGG GAATCCTGAG ACAGTCTAGC TAGTTGGCTA 4260

GCCGTTTCCT CTGGACTAAA AGTATAACCC AAAGCCTCTT GGTTGATGTC ACATATAGCT 4320

TTCACATCAG TTTCTCTTAA ATCTCTTAGC ATCTCATTCC TCCTCAAAAG AAATCTTTGG 4380

CAACCGAGCA AGAATATCTT CTCGCTTAAT GGCCCCTTGA CGTAAGATTT TCACCTTGTC 4440

TCCCGACAAA TTCAAAATAG TTGAATCCTG TCCAGTTAGA AAAGCATCGT CTTCCAGACC 4500

CAGAACCTCT TGGTCAAAAT CCTCTAGAAT TTGATTAAAG GTCACTCCAC TCGCCTGACC 4560

TGAGATATTG GCAGACGGCC CAATCAAGGG ACCTGTCTCT CGAATCAAAT CAAGGGTAAT 4620

GGGATGACTA GGCATCCGAA ATCCAACAGT TGCAAGGCCA GAATTGACCC AATAGGGAAC 4680

TCGGTCATTA GCTTCGAGAA TAATGGTCAA GGGACCTGGT AAAAAGATCT CTACAAGTTT 4740

TTGAAGATAA GTTGGCTGAT TCTTTGAAAA GTACAAGATG TCCTCTAAAG AGGCAACATT 4800

GAGATTGAGC GCCTTGTCTC TACGTCGACG TTTAAGCTGG TAAACATGGT CAACTGCTTT 4860

TTCGTCTAGC GCCTTAGCAA AGAGACCGTA AACTGTCTCT GTAGGCAAAA CGACAGCTCC 4920

ACCATTTTCC AACTCTTGTC TAATCCTGTC CATCATCAAC GACAACCATC CTATCTTGAC 4980

CAAATTGGTC CTTGAGTGTT CGTACTCGCT TTTCAGGAAG ATGTTTCCTA AAAAGTTCAG 5040

GAACACTTTG ACCTTGCTTG TATCCAATTT CAAGGTAAAT CTTACCACCA TCTTTGAGAT 5100

AGTCTTTTGC ATCTTCCGCA ATTCTACGGT AAATAGCTAG GCCATCCTCA TCTGCAAAGA 5160

GAGCTAGATG AGGCTCCGAA TACAAGACAT TCAAGCCTAC CTCTGACTCA TCTTCACGAG 5220

AGATATAGGG TGGATTGGAA ACAATTATAT CATATTTTTC AGAAATTTCT GTAAAACAGT 5280

CAGATTTTTT TAAAAATATT TGAAGATTTT GATTTTTAGC ATTTTCGCTA GCTACATCTA 5340

AAGCATCTTG GGAAATATCT GCTGCCGTCA CTGACCAATC TGGTCTGTTT TTTGCTAGAG 5400

CGAGAGCAAT AGCTCCACTA CCTGTTCCGA TATCTAGGAC CATAAGATTT TTCACAGGAT 5460

TTTCAGCCAG GATAAGCTCC ACCAACTCCT CTGTTTCTGG ACGAGGAATC AAAACCCGTT 5520

CATCCACCTT TAAATGCATT CCATAAAAAT CTGCCTGTCC AATGATGTAC TGAGCTGGCT 5580

TGTGAGCTGC TAGTTGCTGG TAAATATCTT CTACAAATTG TTTTTCTTCC TCTGTTGTCA 5640

CCTCCTGCTG GAGGGCAAAA ATAAAGTCTG TAAAAGATAG ATTTTTCAGA CTACGATAGA 5700

CAAAAGAGAG GCTTTCCGCT TCCTCTCCTT GTCTTATCAA CTCTTCTTCA AAATTTGAAA 5760

ATAATTGAGC TAATTTCATT ATTTGTTTAA TTCTTCTAGT TTTTGTGTTT GGTCATAAAG 5820

CACCAAGGCA TCCACAACTT CGTCCAATTT ACCAGACAAA ATCGTATCTA GTTTTTGGAG 5880

GGTCAAGCCG ATACGGTGGT CTGTGACACG GTTTTGTGGG AAGTTATAAG TTCGGATCCG 5940

TTCTGAACGG TCACCAGTAC CGATTGTCGA CTTACGCTCA GCGTCCTGCT CATCTTGAGC 6000

AATCTGAGCA AAGTGGTCAG CAACACGGGC ACGGATGATT TTCATGGCCT TCTCACGGTT 6060

CTTCTGCTGG GTACGTTCTT CCTGCATCTC AACCTTGATA TTGGTTGGCA AGTGAACGAT 6120

ACGAACGGCA GTCGCAACCT TATTGACGTT CTGTCCACCA GCACCAGAGG CGTGATAGAT 6180

GTCGACACGA AGGTCTTTTG GATCAATGTC GTATTCAACC TCTTCAACTT CTGGCATAAC 6240

AAGAACTGTC GCTGTCGAAG TATGAACACG GCCTTGGCTT TCTGTCACAG GAACACGTTG 6300

CACACGGTGG GCACCTGATT CATACTTAAG CTTAGAGTAT ACAGACTGAC CTGAAACCAT 6360

AGCAACCACT TCTTTAAAAC CACCGACACC ATTCATAGAG GCTTCCATGA CTTCAAAGCG 6420

CCAACCTTGG GCTTCCGCAT ACTTTTGGTA CATAGTTAGC AAATCTCCAG CGAAAAGTGC 6480

CGCTTCGTCT CCACCAGCTG CTCCACGGAT TTCAAGGATG ATATTCTTGT CATCGTTTGG 6540

ATCCTTTGGA AGGAGCAAAA TTTTCAGTTT TTCTTCATAT TCTTCTTTTT CAGCCTTGGC 6600

ATCTTTGAGT TCTTGCTTGG CCAATTCTTC CAAGTCCGCA TCTCCGCCTG ATTCCTTAAT 6660

CATCTCTTCG GCATCGACGA TATTTTGAAG GACTTGTTTA TACTCACGGT AGGCTATTAC 6720

GGTGTCACGA TTGGAAGCTT CTTCTTTTGA AAGCTCCATA AAACGCTTGG TGTCTGAAAC 6780

GACATCAGGG TCACTCAGCA ATTCTCCTAA TTCTTCATAA CGGTCTTCTA CAACTTGTAG 6840

TTGATCATAG ATGTTCATTT TTTCTCCTTA TTTCTCAATT GTTAAATCAT AGATTGCTAC 6900

TACTTCATTC TCGGATATTT CCCCAGTTTC TTTAAATCCA TAACTGAGGT AACAAAATCT 6960

TGCCTGTTCA TTTTCTGGTT CATArGACAA CCAAAGTTTA TTGCTTAAAC CTGCTGGCGC 7020

TGTTCGAACA TAGTCTAGTA CTTTATCCAT AATTGGTTTA AAATATCCTT GATTTTGAAA 7080

ATTCTTATCA ATCATAAAAC GAAATAGTAA ATAATTTCCA CTACTAATTC CGATCTTTTT 7140

ATCATAAGCT ATCATCACAA AACCTATAAT TGCATCATTA TCATAAACTG CCAATGGAGC 7200

TACAAAATCT CCATTTTTAG TGTAGACGTA TGCTTCAGCT AAACTAATTG CGTTGGTTGC 7260

AATGAATTGT TTTTGATATT CCTTGACATC CAAATTTAAA ACATCAAAAT AATTTTCCAT 7320

TGTAACATCT CTTAGTTCAA TTGTCATAGT TTTGCTCCTT GTTAGAGGTT ATCATTGGCG 7380

CAAAATAATG TTTACGGCAA ACTGAGATAT AGGTTTCGTT ACCACCAATC TGGATCTGTT 7440

CTCCATCGTA AACGGGCAGT CCATCCTGTG TTCGCAACAC CATGGTCGCC TTTTTCTTGC 7500

AATACTGACA GATGGTCTTG ATTTCGTCAA TCTTGTCTGC TAAAAGCAAG AGATATTTGG 7560

AACCTTCGAA CAATTCATTG CGAAAGTCAT TTTTCAAGCC AAAAGCCATG ACGGGTATGT 7620

CTAACTCGTC CACAACACGA GCTAGGTCGT AAACATGGTG GCGTTTGAGA AACTGGGCTT 7680

CATCGACCAA AACACAGTAA GGTTTTTCTG GTAGGTCTCG GATATAGCCA AAGATATCCG 7740

TTGTTTCCTC AATCGCAAgG GCAGGGCGTT TCATGCCAAT TCGACTCGAC ACATAGCCAA 7800

CGCCGTCACG CGTATCCAGA GCCGAGGTCA TAATCACAAC ACCTTTTCCT TGCTCCTCGT 7860

AGTTATAGGC CACTTTGAGA ATCTCAATCG TTTTACCAGA GTTCATGGTC CCATAACGAT 7920

AGTACAACTG TGCCATGTTT CTTGCTTCAC GTCCATTTCT AAATTTTTGC TACATTCTAG 7980

TATATCATAA TTTTCTTAAG CTTTAAACGG CAAAATGTGG TAAAATAGAA GAAATCAAAA 8040

ACTAGTGGAG GAAGCTATTA TGCCATTTGT ACGCATCGAT TTATTTGAAG GACGCACGCT 8100

CGAGCAAAAG AAAGCTCTTG CTAAGGAAGT AACGGAAGCA GTTGTCCGCA ACACTGGAGC 8160

CCCTCAATCT GCTGTCCATG TCATCATCAA CGACATGCCA GAAGGAACTT ACTTCCCACA 8220

AGGGGAAATG CGTACTAAAT AAGCTAGCTT AAGCAGAATT GCTTAGGCTT TTTCAATCTC 8280

CAAGTAGCAT TCATTGAAGA AATATCCTAA ATTTGTTACA ATTTGAAAAG AAACTTGGAG 8340

AATTTCCAAG AAAAGAGCTA TTAATTAAAG GAAACATTAT GATTACACGT GAATTTGATA 8400

CCATCGCTGC TATCTCTACT CCACTAGGTG AAGGGGCTAT TGGTATTGTC CGCCTGAGCG 8460

gAACAGACAG TTTTGCTATT GCGCAAAAGA TTTTTAAAGG AAAAGACTTG AACAAGGTTG 8520

CCAGCCACAC TCTCAACTAC GGTCACATTA TTGATCCTCT GACTGGTAAA GTCATGGACG 8580

AGGTTATGGT TGGGGCTATG AAGTCTCCAA AGACCTTCAC TCGTGAGGAT ATTATCGAGA 8640

TTAACACCCA CGGTGGGATT GCGGTGACCA ATGAAATTCT CCAGCTAGCT ATTCGTGAAG 8700

GGGCTCGGTT GGCAGAACCT GGTGAATTTA CCAAACGTGC TTTTTTAAAC GGTCGCGTAG 8760

ACTTGACACA GGCAGAGGCT GTGATGGATA TCATCCGTGC CAAGACTGAC AAGGCCATGA 8820

ACATTGCGGT CAAACAATTA GACGGCTCCC TTTCTGACCT CATTAACAAT ACCCGTCAAG 8880

AAATCCTCAA TACACTTGCC CAAGTTGAGG TCAATATCGA CTATCCTGAG TATGACGATG 8940

TTGAGGAAGC CACTACTGCT GTTGTCCGAG AGAAGACAAT GGAGTTTGAG CAATTACTAA 9000

CCAAACTCCT TAGGACAGCA CGTCGTGGTA AAATCCTTCG TGAAGGAATT TCAACGGCTA 9060

TCATTGGACG TCCCAACGTT GGGAAATCAA GCCTTCTCAA CAACCTCTTG CGTGAGGACA 9120

AGGCTATCGT AACAGATATC GCTGGGACAA CACGAGATGT CATCGAAGAG TACGTCAACA 9180

TCAATGGTGT ACCTCTCAAA TTGATTGATA CAGCCGGTAT TCGTGAAACG GATGATATCG 9240

TTGAACAAAT TGGAGTTGAG CGTTCGAAAA AAGCTCTTAA GGAAGCTGAC CTAGTTCTGC 9300

TAGTACTAAA CGCTAGTGAA CCACTAACCG CCCAAGATCG CCAACTCCTA GAAATCAGTC 9360

AGGAGACTAA TCGCATTATT CTTCTTAACA AAACTGACCT GCCTGAAACG ATTGAAACTT 9420

CGGAACTACC TGAAGATGTC ATCCGCATTT CAGTTCTTAA AAATCAAAAC ATCGATAAAA 9480

TCGAAGAGAG AATCAACAAC CTCTTCTTTG AAAATGCTGG TTTGGTTGAG CAAGATGCTA 9540

CCTACTTGTC AAACGCCCGT CACATTTCCT TGATTGAGAA GGCCGTTGAA AGCCTACAAG 9600

CTGTTAACCA AGGTCTTGAA CTAGGGATGC CAGTTGACTT GCTTCAAGTT GACTTGACCC 9660

GTACTTGGGA AATTCTAGGA GAAATCACTG GAGATGCTGC TCCAGATGAA CTCATCACCC 9720

AACTCTTTAG CCAATTCTGT TTAGGAAAAT AAGAAAAATC CATGATCCTT CATTCGGTCA 9780

TGGATTTTAG GTTCTATAAT ATTTGTAGTG GGTAAATCCA CTATAGATAT TATGGAGCCT 9840

ATTTTATTGT AGAAAAAAAG TCCCATATGA CCTATAATGA AAAGCGACAA AACAACTCAT 9900

TAGAAAGAAT CATATGGAAC AATTACATTT TATCACAAAA TTACTAGACA TTAAAGACCC 9960

TAATATCCAG ATTTTAGACA TCATCAATAA GGATACACAC AAGGAAATCA TCGCCAAACT 10020

GGACTACGAC GCCCCATCTT GCCCTGAGTG CGGAAACCAA TTGAAGAAAT ATGACTTTCA 10080

AAAAACCTTC TAAAATTCCT TATCTTGAAA CGACTGGTAT GCCCACTAGA ATTCTCCTTA 10140

GAAAGCGTCG ATTCAAGTGC TATCACTGTT CAAAAATGAT GGTCGCTGAA ACTTCTATCG 10200

TCAAGAAGAA TCACCAAATC CCTCGTATCA TCAACCAAAA GATTGCTCAA AAGTTAATTG 10260

AAAAGATTTC TATGACTGAT ATTGCCCATC AGCTTTCCAT CTCAACTTCA ACTGTTATTC 10320

GTAAGCTCAA TGACTTTCAC TTTAAACATG ATTTTTCTTG TCTTCCTGAG ATTATGTCTT 10380

GGGATGAGTA TGCTTTTACA AAAGGGAAGA T 10411

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2393 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

GTTTTGGGTT CTGGAAATTA TCAGATGGTT GGAAAAGCCG TCCACATCAA GATAGTGTTC 60

GGAGATTTAA GTTTAAATTG AAGAAACTAA CACAGAGGAA ATGGAGTATA GACCTAACAA 120

GACGTATTGA GCAACTGAAT TTGTCTATTC GAGGATGGAT AAACTATTGC TCATTGGGAA 180

ATATGAAAAG TATAGTCGCC AGCATAGATG AGCGCTTGCG TACTCGCCTA CGAGTGATTA 240

TCTGGAAGCA ATGGAAGAAG AAATCGAGAC GATTATGGGG ATTGCTTAAG TTAGGAGTTC 300

CTAAATGGAT AGCAGATAAG GTATCTGGCT GGGGCGACCA TTATCAATTA GTAGCTCAGA 360

AGTCGGTACT TAAACGTGCT ATATCAAAAC CAGTCCTGGA AAAACGTGGA CTGGTTTCGT 420

GTTTGGATTA TTACCTTGAA CGACATGCGT TAAAAGTTAG TTGAACCGCC GTATGCCAAA 480

CGGCACGTAC GGTGGTGTGA GAGGGGCTAG AGATTATCCC CTACTCGATT AACTCCCCTG 540

AAATTTATTT TAATTATGCA AATTTCACGT ATTTTTGATG CTGAGACGAC GATCCTGGGA 600

ACTTTTCAGA TATTTTTTTG ACTATCTAAA TCTATCATTA GAAAAGCTTA GAGCGCCAAA 660

GGATTTGAGC GTTTTTCTGA TTTTTAAGAC TTTTTCCAGT CTCTTTTTCG ATTGAAGATG 720

TAATTATTCT ACTAACTAAC TAACTTCTTA GTACTAGCCA ACAACGATAA TCATAATTCC 780

TCCTAAAATT AGGAATAATA AAGGCAATAG TTTTTGTTTT TTCATGTAAA AAACCTCACT 840

TTTGTTTTCT GCTATTTTAT GCTAAAATAT TAAAAATCAA ATTTAATTCC AAAGTTTGTA 900

ACTAAAGGGG GAGCGCTACA TGTCTAATTC ATTTGTCAAG TTGTTAGTCT CTCAATTATT 960

TGCAAATTTA GCAGATATTT TCTTTAGAGT AACAATCATT GCTAACATAT ACATTATTTC 1020

AAAATCAGTA ATTGCCACAT CACTAGTTCC TATCTTAATA GGAATATCCT CTTTTGTTGC 1080

GAGTCTTTTA GTTCCGTTGG TTACTAAAAG GTTAGCGCTA AATAGGGTTT TATCTTTATC 1140

TCAATTTGGA AAGACTATAT TATTGGCGAT ACTGGTAGGA ATGTTTACCG TAATGCAATC 1200

CGTAGCGCCT TTGGTGACCT ATCTATTTGT TGTTGCAATT TCCATACTAG ATGGTTTTGC 1260

AGCACCCGTT TCCTATGCTA TTGTGCCACG CTATGCGACC GATTTGGGTA AGGCTAATTC 1320

AGCCTTATCA ATGACTGGTG AAGCTGTTCA ATTGATAGGT TGGGGATTAG GTGGACTCTT 1380

GTTTGCAACA ATTGGTCTGT TACCTACCAC GTGTATCAAT TTAGTCTTGT ATATCATTTC 1440

TAGCTTTCTG ATGTTATTTC TTCCTAACGC TGAAGTGGAG GTGTTAGAGT CAGAAACTAA 1500

TCTTGAAATT TTGCTCAAAG GTTGGAAGTT AGTTGCTAGA AATCCTAGAT TAAGACTTTT 1560

TGTATCAGCA AATTTATTGG AAATTTTTTC AAATACGATT TGGGTTTCTT CCATTATACT 1620

TGTTTTTGTA ACGGAGTTAT TAAATAAAAC GGAAAGTTAC TGGGGATATT CTAATACAGC 1680

ATACTCTATT GGTATTATAA TTAGTGGCTT AATTGCTTTT AGGCTATCTG AAAAGTTCCT 1740

TGCTGCTAAA TGGGAACCCC AATTATTCAC CCCAAATCTA AAAACCATCC AGAATCCTTG 1800

CCTTAGCTTA GATCCTGGAT GGTTTCTTTT TTCACCCAAT GGGTGTTTTT TACTAGACAA 1860

AAAAGAGTTT CCCCTTTATG GTATAAGTGT AGAAAAAAAC ACAAAAAGAA AGGAAACTCA 1920

CATGAACAGT TTACCAAATC ATCACTTCCA AAACAAGTCT TTTTACCAAC TATCTTTCGA 1980

TGGAGGTCAT TTAACCCAGT ATGGTGGTCT TATCTTTTTT CAGGAACTTT TTTCCCAGTT 2040

GAAACTAAAA GAGCGGATTT CTAAGTATTT AGTAACGAAT GACCAACGCC GCTACTGTCG 2100

TTATTCGGAT TCAGATATCC TTGTCCAGTT CCTCTTTCAA CTGTTAACAG GTTATGGAAC 2160

GGACTATGCT TGTAAAGAAT TGTCAGCTGA TGCCTACTTT CCAAAATTGT TGGAAGGAGG 2220

GCAGCTTGCT TCACAGCCAA CCTTATCCCG TTTTCTTTCC AGAACTGACG AGGAAACAGT 2280

CCATAGTTTG CGATGCCTCA ACCTTGAATT GGTCGAATTC TTTTTACAGT TTCACCAGCT 2340

AAACCAACTC ATTGTAGATA ACGATTCTAC CCATTTCACA ACTTATGGCA AGC 2393

(2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4762 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

TTTGTATCTT TTTAGGTCTC TTTCAATCCA AACCCTTTAA ACTATACGTC ATTTCGGTTC 60

CTGCAAGTCT TGTGGTAATT TTAGGTTTGA TTTTACTTTT CTTTTCACAA GAGCCTCTGC 120

ACGCTTCTTA TTTGATGGTC GTCTTCCCTG TTTTCCTACT TTTATTGGTA ACCAATATTA 180

AGAGTCAACA GAGGGGGCGT AGTGCTAGAA GAAGCCGAAG AGAAACGCCA TTATGCCTAT 240

GGAGTCGTTT CTTCAAAGGA AATCTATATC TGCTAGTTTT TGGGTTTGTC TATCTTTTGT 300

CTGTTCCTTT TTTGATGAAG TTTGTCCTTT ATCCAGTACC TTATCAAGAA CGTAATCGTC 360

TTGCTGATTT GGTAAAAGAG GAGACAAATA CGGAAGATGC TATCTCATGC ATGGGATGAT 420

ACTGCGACTC TTTATCGTAA GAGTGAGCGC TTGTCCCATC GGCGATTTTG TCCCCGTTGC 480

ACTATACAGC AACTGAGGAA AATCGTAATA AGTTACTTAA TGACTTGAAA GAAAAACAAC 540

CTAAGGTGAT TGTGGTAAAT GATAAGGTGG TAGTCTGGTC TGAAGTGGAA ACACTCTTAA 600

AAGAAAATTA CCAACAAGTA AAGACTGATT ACTCAGAGTT TAAAGTCTAT AAAATTAAAT 660

AACCAAATCA ATATCTTGTG TATTTTTAAA AATTTTAGGA TTTTTAACAC AAGATATTGA 720

TTTTTCTTTT TAGAGTGGTA TAATACTTTT TAGAAAGAAC ATTTTAGAAA AGAGCATGCA 780

TATGATTGCA CTAGAAGAAA AAATTACAAT TTTGCCAACT CTCTTCGTCG AGAAACGAGA 840

TGGGAGACGT GTTGTATTTG ATGTGGACAA GATTGACAAG GCTCTCCACA AGGCGGCTGA 900

CAAGGTTATG GATGTGACAC CCCTGGTTGA AAAATGCCTC AATGATCTGA CTGAGCGAAT 960

TATTACAGAA ATTCATAGTC GCTTTCCACA GGGAATTAAG ATTTACGAAA TTCAAAATAT 1020

CGTAGAACAT GAACTCCTTG AAGCCAAAGA ATATGCGCTG GCTGAGGAGT ATATTACTTA 1080

TCGGACACAG AGGGATTTTG AGCGCTCAAA AGCGACGGAT ATCAACTTTA GTATTCATAA 1140

ACTTCTCAAC AAAGACCAGA CAGTTGTCAA TGAAAACGCT AATAAAGACA GTGATGTCTT 1200

TAACACTCAG CGTGATTTGA CAGCAGGGAT TGTTGGGAAA TCAATCGGAC TGCAAATGCT 1260

TCCTAAGCAC GTAGCCAATG CCCACCAAAA GGGGGATATC CACTATCACG ATTTGGACTA 1320

CAGTCCCTAT ACCCCTATGA CCAACTGCTG TTTGATTGAT TTTAAGGGTA TGTTGGAAAA 1380

TGGTTTTAAG ATTGGAAATG CAGAGGTAGA GAGTCCCAAG TCTATCCAGA CTGCGACAGC 1440

ACAGATTTCT CAAATCATTG CCAACGTTGC TTCTAGCCAG TACGGTGGCT GTTCAGCTGA 1500

CCGTATCGAT GAAATTTTGG CGCCTTATGC AGAGAAGAAT TATCAAAAAC ATCTCAAAGA 1560

TGCAGAAGAG TGGGTATTGC CTGAAAAACA GGAAGATTAC GCTTGGAAGA AAGCGCAAAA 1620

GGACATCTAC GATGCCATGC AATCTCTTGA GTATGAAATC AATACTCTCT TCACTTCAAA 1680

TGGACAAACA CCTTTTACTT CGTTAGGTTT TGGTCTGGGA ACCAGTCGTT TTGAACGAGA 1740

AATTCAAAAA GCTATTTTAA ACATTCGCAT CAAGGGTCTT GGTTCAGAAC ACCGTACGGC 1800

TATCTTTCCT AAACTTATCT TTACGCTTAA AAGAGGCCTC AACTTAGAGG AAGGAACTCC 1860

CAACTATGAC ATCAAGCAGT TGGCTCTAGA GTGTGCAACC AAGCGGATGT ATCCAGACGT 1920

CTTGTCTTAT GATAAGATTG TTGATTTGAC AGGTTCTTTC AAGGTGCCTA TGGGCTGCCG 1980

TTCTTTCCTT CAAGGGTGGA AGGATGAAAA TGGTGTAGAA GTCAATTCAG GTCGCATGAA 2040

TCTGGGTGTT GTGACGGTTA ATCTGCCTCG TATTGCTCTT GAGTCTGAAG GTGATATGAA 2100

TAAGTTCTGG GAAATCTTCA ACGAGCGAAT GAATATCGCA GAAGATGCTC TTGTTTACCG 2160

TGTCGAACGC ACTAAAGAGG CGACACCAGC GAATGCTCCT ATTCTTTATC AGTACGGTGC 2220

TTTTGGCCAT CGTCTAGGTA AAGAAGAAAG TGTTGACCAG CTCTTTAAGA ATCGTCGTGC 2280

GACCGTTTCG CTGGGCTATA TCGGCTTGTA TGAAGTAGCG ACAGTTTTCT TTGGTAACAG 2340

CTGGGAAAGT AATCCAGATG CTAAGGAATT CACGCTAGAC ATCATTCACG ATATGAAACG 2400

CCGTGTAGAA GAGTGGTCAG ACCAATATGG CTACCATTTC TCTATCTACT CAACACCATC 2460

CGAAAGTCTG ACAGACCGTT TCTGCCGACT AGATATAGAC AAGTTTGGCT CTATTCCTGA 2520

TATCACAGAC AAGGAATACT ACACCAACTC TTTCCACTAC GATGTTCGTA AAAATCCAAC 2580

ACCGTTTGAA AAATTGGACT TTGAGAAAGT CTATCCGGAA GCAGGTGCGT CAGGTGGTTT 2640

CATCCATTAT TGTGAGTATC CAGTCCTTCA GCAAAATCCA AAGGCCTTGG AAGCTGTCTG 2700

GGATTATGCT TATGACCGTG TAGGCTATCT AGGCACCAAT ACTCCGATTG ACCGTTGCTA 2760

CAAGTGTGAC TTTGAAGGGG ATTTTGAACC AACTGAGAGA GGGTTTGCTT GTCCAAACTG 2820

TGGCAATAGC GACCCTAAAA CAGTAGATGT GGTGAAACGA ACTTGTGGCT ACCTAGGTAA 2880

TCCTCAAGCA AGACCGATGG TCAACGGGCG TCACAAGGAA ATCGCTGCGC GTGTCAAACA 2940

TATGAATGGT TCAACGATTA AAATAGCTGG GCATCAAGTA ACAAATTAGA AAGAAATGAA 3000

ATGGGAAAAT ATCAACTAGA CGATAAGGGG CGCGCACAAG TGACCCGTTA TCACGAGAAA 3060

CACTCTAAAG GTGGAGCTGG TAAGAAAGAA CGCTTGCTTA GCTTCAGAGA ACAATTTTTA 3120

AACAAGAACA AGAAAAAATA AAAGTGAGAG CCAGCTCTCG CTTTTCTCAT AGTGGGAGGT 3180

AAGGATGGAA TTACGCAGAC CAAGATTAGC GGATAAGAAA GCTGTTTTAG ATATGATGAC 3240

AGAGTTTGAA AAATTTCAGT CGCCTCACGA CGGCGGTTTC TGGGATACAG AGAACTTTGT 3300

GTATGAAGAC TGGTTAGAAA GCAATCAGGA ACAGGAAATG GGGATTAATC TGCCTGAAGG 3360

ATGGGTTTCT GCAATTCAGT TAGTGGCTTT TTCTGAGAAA GGTCAAGCAG TTGGATTTCT 3420

TAATCTCCGG TTGCGCCTCA GTAACTTTCT ACTAGAAGAA GGTGGCCACA TTGGCTACTC 3480

CATTCGTCCA TCTGAAAGAG GCAAGGGTTA TGCAAAAGAG ACTCTCCGTC AGGGCTTGCA 3540

AGTTGCTAAG GAAAAGAACA TCAAGAAAGC TCTGGTGACC TGTAGTGTGA ATAATCCTGC 3600

TAGCAGAGCA GTCATTCTAG CAAATGGTGG AATATTTGAG GATGCTCGCA ATGGAGTCGA 3660

GCGTTATTGG ATAGAGGTAG CGAATGAATA ATCCAAAACC ACAAGAATGG AAAAGCGAGG 3720

AACTTAGTCA AGGTCGTATC ATTGACTACA AGGCCTTTAA CTTTGTGGAC GGCGAAGGCG 3780

TGCGCAACTC TCTCTATGTA TCAGGCTGTA TGTTTCACTG CGAGGGATGT TATAATGTTG 3840

CGACTTGGTC TTTTAATGCT GGCATTCCCT ATACAGCAGA ATTAGAAGAG CAGATTATGG 3900

CAGACCTTGC CCAACCCTAT GTTCAAGGCT TGACTTTGCT GGGAGGGGAG CCTTTTCTCA 3960

ATACTGGGAT TCTCTTGCCA CTTGTTAAGC GGATTCGGAA GGAATTGCCA GACAAGGACA 4020

TCTGGTCCTG GACCGGCTAC ACTTGGGAAG AAATGATGTT GGAAACTCCA GATAAACTGG 4080

AATTCTTGTC ACTGATTGAC ATTCTTGTCG ATGGAAGATA TGATCGAACT AAGAGAAATC 4140

TTATGCTCCA GTTTCGAGGT TCATCTAACC AACGAATTAT CGATGTGCAA AAATCGCTCA 4200

AAAGTGGGCA AGTAGTGATT TGGGACAAGC TCAATGACGG AAAAGAAAGC TATGAACAGG 4260

TGAAGAGAGA ATGAAGAAAA AGGACTTAGT AGACCAACTA GTCTCAGAGA TCGAGACGGG 4320

GAAAGTCAGG ACACTGGGAA TATACGGTCA TGGAGCTTCA GGTAAATCAA CCTTTGCACA 4380

GGAATTGTAC CAAGCTTTAG ATTCTACTAC AGTAAATTTG CTAGAGACAG ATCCTTATAT 4440

CACCTCAGGA CGCCATCTGG TAGTACCCAA GGACGCGCCG AATCAAAAGG TGACAGCCAG 4500

TCTGCCAGTG GCGCATGAAC TGGAGAGTTT GCAGAGAGAT ATCCTTGCTT GCAGGCGGGT 4560

ATGGATGTCT TGACAATTGA AGAACCTTGG AAGGCTAGTG AGGTCTTGTC TGGAGCCAAA 4620

CCAATTTTGA TTGTCGAAGG GATGTCTGTT GGCTTTCTAC CCAAGGAACT CTTTGAAAAA 4680

ACCATCTGTT TCTACACGGA TGAGGAGACC GAATTAAAGC GACGCCTTGC TAGAGATACG 4740

ACTGTGAGAA ATCGCGATGC GG 4762

(2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3832 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

GATGCAGGTT TCGACCCACA TATTCCAGAA AATTACTTTA AAGATGATGA TGTTAATCAG 60

GTACCTTGTC TTTGTTGGTC TTCATCTGCA GCCCTCTTTT TCAGTAATTG GGTAGACCAT 120

GCGGTCTATC AGGAGACGCC TTTTGATTGG AGAAAGATAG AAGATGATGC ATCTGCATAT 180

GGGTATTTAT AAGAGGAATT ATGACATATT TAGACGCTTT TAAATCAGGT ACCTTGGTTT 240

TACCGAGTGC CCTGCTCTTG CATTTTAAGG AACTCTTTCC TTCTAGCGAC GATTTTCTGG 300

TTTGGCAATT TTTCTATTTG CAAAATACGA CAGGCTTAGA AGAAATGTCG CCAAGCCAGA 360

TTGCTGAAAG GATTGGCAAG GAAATTTCGG ATGTCAACCA GTCCATTTCT AATCTGACGG 420

AAAGGGGACT GCTCCAGTAT CGTACTATCG AATTAAATGG CGAAATTGAA TTGCTCTTTG 480

ATGCTAGTTT GGCCTTGGAA CGTTTGGATG ACCTGTTTGG AGCAGTTCAT TCAAGTTCAG 540

ACCAGCTAAC ACCTCAAAAC CAGCTCAAGG ATTTGGTGGA AACCTTCCAG CAGGAGTTGG 600

GACGATTGTT GACGCCTTTT GAGATTGAGG ATTTGACCAA GACACTAAAG GAAGATGGAA 660

CCAGTGCTGA CTTGATTAAG GAGGCTCTTC GTGAAGCTGT TTTGAATGGA AAACCAAACT 720

GGAAGTACAT TCAGGCGATT TTGAGAAACT GGCGCCATGA AGGAATCAAG AGTGTGGCTC 780

AAATTGAGGC CAAGAGAGCA GAAAGAGAAG CAAGCAATCC TCAGTTGACA CAGGTATCTG 840

CAGATTTCAT AAATGCCATG GATCTCTGGA AGGATTAATC CATGCAAGTA GGCTTGAAAT 900

CCGAGTAAGA TTTGCAAGCT GTGTATAATT GTGATAGAAT AAATAGAAAA TAAATTGAAA 960

AAAGAGGTAT GTGAAATGTC ACGTAAACCA TTTATCGCTG GTAACTGGAA AATGAACAAA 1020

AATCCAGAAG AAGCTAAAGC ATTCGTTGAA GCAGTTGCAT CAAAACTTCC TTCATCAGAT 1080

CTTGTTGAAG CAGGTATCGC TGCTCCAGCT CTTGATTTGA CAACTGTTCT TGCTGTTGCA 1140

AAAGGCTCAA ACCTTAAAGT TGCTGCTCAA AACTGCTACT TTGAAAATGC AGGTGCTTTC 1200

ACTGGTGAAA CTAGCCCACA AGTTTTGAAA GAAATCGGTA CTGACTACGT TGTTATCGGT 1260

CACTCAGAAC GCCGTGACTA CTTCCATGAA ACTGATGAAG ATATCAACAA AAAAGCAAAA 1320

GCAATCTTTG CGAACGGTAT GCTTCCAATC ATCTGTTGTG GTGAATCACT TGAAACTTAC 1380

GAAGCTGGTA AAGCTGCTGA ATTCGTAGGT GCTCAAGTAT CTGCTGCATT GGCTGGATTG 1440

ACTGCTGAAC AAGTTGCTGC CTCAGTTATC GCTTATGAGC CAATCTGGGC TATCGGTACT 1500

GGTAAATCAG CTTCACAAGA CGATGCACAA AAAATGTGTA AAGTTGTTCG TGACGTTGTA 1560

GCTGCTGACT TTGGTCAAGA AGTCGCAGAC AAAGTTCGTG TTCAATACGG TGGTTCTGTT 1620

AAACCTGAAA ATGTTGCTTC ATACATGGCT TGCCCAGACG TTGACGGTGC CCTTGTAGGT 1680

GGTGCGTCAC TTGAAGCTGA AAGCTTCTTG GCTTTGCTTG ACTTTGTAAA ATAATCAGTA 1740

AGTAGCAAAA GCTAGGTGGA ACAGCATTCA GATGTCTGTT ACATTTTTTA TAGGAGAGAA 1800

AGATTGAAAA CAAAAATTGG ATTAGCAAGT ATCTGTTTAC TAGGCTTGGC AACTAGTCAT 1860

GTCGCTGCAA ATGAAACTGA AGTAGCAAAA ACTTCGCAGG ATACAACGAC AGCTTCAAGT 1920

AGTTCAGAGC AAAATCAGTC TTCTAATAAA ACGCAAACGA GCGCAGAAGT ACAGACTAAT 1980

GCTGCTGCCC ACTGGGATGG GGATTATTAT GTAAAGGATG ATGCTTCTAA AGCTCAAAGT 2040

GAATGGATTT TTGACAACTA CTATAAGGCT TGGTTTTATA TTAATTCAGA TGGTCGTTAC 2100

TCGCAGAATG AATGGCATGG AAATTACTAC CTGAAATCAG GTGGATATAT GGCCCAAAAC 2160

GAGTGGATCT ATGACAGTAA TTACAAGAGT TGGTTTTATC TCAAGTCAGA TGGGGCTTAT 2220

GCTCATCAAG AATGGCAATT GATTGGAAAT AAGTGGTACT ACTTCAAGAA GTGGGGTTAC 2280

ATGGCTAAAA GCCAATGGCA AGGAAGTTAT TTCTTGAATG GTCAAGGAGC TATGATGCAA 2340

AATGAATGGC TCTATGATCC AGCCTATTCT GCTTATTTTT ATCTAAAATC CGATGGAACT 2400

TATGCTAACC AAGAGTGGCA AAAAGTGGGC GGCAAATGGT ACTATTTCAA GAAGTGGGGC 2460

TATATGGCTC GGAATGAGTG GCAAGGCAAC TACTATTTGA CTGGAAGTGG TGCCATGGCG 2520

ACTGACGAAG TGATTATGGA TGGTACTCGC TATATCTTTG CGGCCTCTGG TGAGCTCAAA 2580

GAAAAAAAAG ATTTGAATGT CGGCTGGGTT CACAGAGATG GTAAGCGCTA TTTCTTTAAT 2640

AATAGAGAAG AACAAGTGGG AACCGAACAT GCTAAGAAAG TCATTGATAT TAGTGAGCAC 2700

AATGGTCGTA TCAATGATTG GAAAAAGGTT ATTGATGAGA ACGAAGTGGA TGGTGTCATT 2760

GTTCGTCTAG GTTATAGCGG TAAAGAAGAC AAGGAATTGG CGCATAACAT TAAGGAGTTA 2820

AACCGTCTGG GAATTCCTTA TGGTGTCTAT CTCTATACCT ATGCTGAAAA TGAGACCCAT 2880

GCTGAGAGTG ACGCTAAACA GACCATTGAA CTTATAAAGA AATACAATAT GAACCTGTCT 2940

TACCCTATCT ATTATGATGT TGAGAATTGG GAATATGTAA ATAAGAGCAA GAGAGCTCCA 3000

AGTGATACAG GCACTTGGGT TAAAATCATC AACAAGTACA TGGACACGAT GAAGCAGGCG 3060

GGTTATCAAA ATGTGTATGT CTATAGCTAT CGTAGTTTAT TACAGACGCG TTTAAAACAC 3120

CCAGATATTT TAAAACATGT AAACTGGGTA GCGGCCTATA CGAATGCTTT AGAATGGGAA 3180

AACCCTCATT ATTCAGGAAA AAAAGGTTGG CAATATACCT CTTCTGAATA CATGAAAGGA 3240

ATCCAAGGGC GCGTAGATGT CAGCGTTTGG TATTAAGCGA TGATTTGAAA GAGGGATGTG 3300

ATAGTAGCAC CCTCTTTTTC TTTGTTTTAT GATAGTTCAT CCTCGAGTAA ATTCAAGTTC 3360

TTGCTCGGAA ATGAAGCTTA TATAGTAGAT TGAATATAGA CAAATACCTT GTGATTGGTA 3420

AAACATTTTA GAAATTCATT TACCTTTCCT AATCGACTTG GTTTCATCTT ATTTCAATCT 3480

ATTATAGTAT TGGGGAATTT CTTCAAACCA CATCAGCTTG GTCAGTTCTA CCTGCGACCT 3540

CAAAACTTGT GCTTTGGTCA AGCTGGGTTT AGTTTCCTAG TTTGCTGATG GATTTCCATT 3600

GACTATAAGC ATCCAACCCT CTTTTTGTCT TCTAAAGAAT TCTTAAATTA TCAGTCTATT 3660

GCAACTTTTC TCATATAAGT TCTTTGTCTT GCTATTGGTT TTCCTTAGTA GTATACTAAG 3720

GTAGTAATCA TTAAGAAGTG GTTACAAAAA ATAATGAATG AGGTAAAGAA AATGGTAGAA 3780

TTGAAAAAAG AAGCAGTAAA AGACGTAACA TCATTGACAA AAGCAGCGCC GG 3832

(2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10690 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

TGAAAAAATC CTCATGAACC TGGCGCCAAT AGACAAGTGT CTTGTTTCCC TCACCTTCCT 60

TATAGGCATG GTCAGCTGAC ACTCGATTGA AGGGTTTAAC AGAAACCTTT GTAATTTCGA 120

CAATGCAGAC AGCCTGATTT TGACTATCTA AAATGACATC GAAGGTCCCT ACTTGGGGAA 180

GTGGTTCGTC TTCTAGCACA TAGAGGTCAT AGGCTGATGC TGTTGCTGTC TTTTCTCCTT 240

TAAACACCAA ATCCGCTAAA AGGTCTGGTT CAACTCCAAA AGCCCAGGCA TCGATTTCAT 300

CTCCGATCAA AGGATTGATT TGCTTGTATT TATTCCACAT TTCTTGCGGT ATCATGGGTG 360

CTCCTTTGTA ATTTTTTACT TTCTTCTTTT ATGTGTTTAA GATGATCTGG ATGGTCAATC 420

TCTAAATCAA AAATCTCTGG AATAGAACTG TAGTGGATAA TGCACTTGAT ACCCAACTGA 480

TTCATTTTTT GTATGAAAGA AGTATTCAGA TAGCCTGCTA CAGCAAAATC AATCTTGTTC 540

TTTCTTGCTT TATCCTGCAT ATCTCTTAGC ATATCTAACA TTATTGGACT TTCCATATCA 600

TGCCATTGAC TGTTTCTCAT AGTCGCAAAA ACAAAGGAAG TCAAATCATT CATTCCAACT 660

ACAATCTTTG AAATGCCCGT TTCCAGTATA CTAGATAAGT CAAAATACGC TGACGGTAAT 720

TCAATCATCG TTCCGACTTT CCCAGTAAAA CCCTGCTGAC GCAATACTGT AATAGCTTGT 780

TTTAATTGGT CGGCATCATT GACAAAAGGA AAGATAACAG ATAGATTGGG GTTGGTTTGA 840

TAAACTTCTG TAACGACATG TGCTTCAGCC TGAAATTCAT CCAAACACGC CAGTAAACGC 900

CTAGTTCCTC TATAGCCAAA CAAGGGATGC CCTTCGTCAA AAAACTCTTT AGTCCCCACT 960

AAACAATTGG CTTCTGTATT CGTTAATTCA GTAAAACGAT ACCAAACTTC CTTACCTAAG 1020

TAAAAGGAGC AAATAGTATC AAGATAATCT TTCACAAATT CCTGACAACT TTGTAATAGT 1080

ATATTTTGAT TGAGCTCTCT CAATAAGTAT TCCCCACGAA TCATGCCGAC GTGGTGAAAT 1140

AGTTGAGGAT AAATTTTTTC AAGAATTTTT TCGCCACTAA GGGCAAGTTG ATTTCTCATC 1200

ATTCACCTTC CAATTCATGT AAGAAGTCTT GTCCAGTTCT GGAAATCCTA ATAATTCAGA 1260

CTTAACCTTC AAGACTAATG GCGATGCATT TTCTTCTGTA ATCTCTTGAA TATCCATCCA 1320

AATATATCCA AGTGAATCAT TCGCACCATC AGACACAGCT TCCGAAATCG TAACTTGAGG 1380

TGCACTCTCA TTCATTTCAA CATCATACAA GGCTATGACA TGGTGAACCA TAAAATTTTT 1440

TAACTCTTCC CTGACGAAAA CATCGTAGAT TCGAGGATTA GAGTAGCTTC TAACAGTAAA 1500

TCCCGTCTCT TCCATAACTT CTCTAGTCAG CGTTTCCGTC AGTCCTTCAC CAAGTTGCTG 1560

ACTGCCTCCA GGTAGATCAT ACCGATGTTG ATAAGGGCCT CTCGTTTTTT CAATGCAAAG 1620

TAACTTTCCA TTTTCAAAGC AAACACAGTA GACCCCAAAG TGATTTTTGA TTTCCATCCA 1680

ACTCCTCCTA CTTCAAAGAC CAGCCACCAT CTATTGTCAA GATTTGTCCT TGCATGGCGC 1740

TCGCTTTTCC ACTTGCTAAA AAAAGACTAA GCTCTGCTAT TTCCTCTGGC TCAATCCAGC 1800

GCTTGATTGG GGTTTCACTA GCCACCCAGT CAGCCAAACC ACCTGGTTCA AAATCCGCAG 1860

CGGTCATAGC TGTCTTGACT GCTCCTGGAG CGATACCAAA GACCTGAATC CCAGCTTCAG 1920

CATAGTCTAG AGCCAACTGC TTGGTGAAGC CAGCCAAGGC ATGCTTGGAT GAAGTATAGG 1980

CGTGACCACC TCCACCTGCT AGGCTAGAAG CAATGGAACA CATATTGATG ATGATTCCCT 2040

TTTTATTTTC CAGCATTTGT GTCAAATAAT ACCGAGTCAA CTCTACTGGA ATAATGTAGT 2100

TGATTTCAAA AATCTCTTGA ATGTCCTGCG CCGTTTGTTC CAACAGTGGT TTGTAATCAT 2160

CCAAAACTCC AGCAGTATTA CACAAAACAT CCACCTGAGG GCACCAGTCA AAAATAGGTT 2220

CCAAGTCCAA GGTCAAATCT CTCTGTAAAA AGCGAAAATC ACCCTCTAAG AGTGGCTTTT 2280

CACCTTGGTC AACTCCATAA ACTTGATAGC CCTTCTCTAA AAAGAGGCGA GCTTGAGCCA 2340

ATCCGATCCC TGAACTCACT CCTGTAATGA GTACACGTTT AGTCATGCAC TTCTACCCAA 2400

TCCGTTGCCA AAACATCACA AACTGTCGGG CTCCACATGG AAAAACCTTC TCCTTCGCCA 2460

GAAACGTTGA TTAGGAAATA AGGTGTCATT TCAAGTGCAA GCCCATTTTG CTCGATGGTA 2520

TCAAAGAGTT GGACATAGTT TTCCGCACCT CCCCAACCAG TTCGTACATA TTTTCTCTTA 2580

GCCTTTAACC CAGGCAGGAT CTCTTCAAAT GTCATGTTTT TCTCCTTTAA TTCTACATTC 2640

TTCATTTAAT TATAGCAAAA AACCGCTTTA TACGGCTTTT TGAATGTGAG TTATTCAAAC 2700

CTGCTACTAC TTACGGCAAA TTATTCCCTG CAGCAAGATA AATTTCATAC CATTCTTTTC 2760

TTGTTAAGCT AAAGTTTGCC GCTCGGCTAA CTTCTCTCAA GTGCTTAGGA TTTGTTGTAC 2820

CTACGACTGC CTGCATTTTT GCTGGATAAC GCAATATCCA AGAAATGGCA ATAGTTGAAG 2880

AGGTTACTCC ATATTTAATA GCTAAACGAT CAAGTACTTG ATTTAAAGCT TGAAATTTCT 2940

CATTTCCAAC AAAATTCCCT TTAAAATACC CGAATTGTAA GACAGACCAT GCTTGAATGA 3000

CCACATCGTG TAATTGGCAA TATTCAAAAA TGCTGCCATC TCGCATAGCT GCTTGACTAT 3060

CTTCCATATT AACATGAAAA GCTGATTCAA ATCCTGGAGT AAAAGCCGCA CTCAATTGTA 3120

GCTGATTAAC AGCTAACGGC TGCTTGACAT CTTTTTTAAG CAACTCCATC ATCATAGGAT 3180

TTTGATTAGA AACTCCAAAA TCTCGAACTT TACCTTGTTT ATAAAGGAGA TTAAAGGCTT 3240

CTGCTACTTG GTCAGATTCC ATCAAAGCAT CTGGTCGATG AAGGAGCAAG CTATCTAGAT 3300

GATCAATCTT CAATCTTTGC AAAATACCGT CTACTGATTT TATAATATAG TCCTTAGAAA 3360

AATCAAAATA GGTAAATTCT TCAATGCGAA TGCCACATTT GGACTGAATC CACATCTTTT 3420

CTCTTAAATC TGGACGATTT TTTAGGACAA GACCTAACAG TTCTTCACAA CGACCACGAC 3480

CATAAATATC AGCCAAGTCG AAGGCATTGA TTCCAACAGA AAGTGCTGTT TCTACAAGCT 3540

CTTCAACTTC TTTTACAGAT TTATCTTTTA TTCTCATCAT TCCGAGAACA ATTTCTGATA 3600

ATTCTTTGTC ATCTTGACCA AGAGTTATGT ATCTCATCAA ATTTTTCTCC TTTAATTTCT 3660

AACATTCTTC CCTTCATTAT AACAAAAAAC CGCTTTGCAA CGACTTTTTG ACTATACTTC 3720

ACTCCATTTT ATCTTCTTAA ACCCACGGAA CAAGACAAAG ATTCCAATAA AGAGGACAGC 3780

TAAAGGAATA ACTTTTGTAA GGAAAACATT TGAAATTCCC ATCCACTCAT AATAACGGAG 3840

CAGAGAACCC ACCACAAGAT GGGCAATAAT CATACTGACA AATGGACGAA AGACCGCTTC 3900

TTTCCAATTC CAAATACCGA TAACTAGCGA AATCGTAAAG ACAGACAAAC TATCCCAGGG 3960

AGCCGGAATA TAAAAGGCTC CTTCTTGTAT GAAGCTTGCC ATTCCTACAT ATCCTAAAAC 4020

AACTAGAAGA ACTATAGTCC CAACAACAAT GTAAGTGCCA ATTTTCATTT TAGGAGAATC 4080

TTGGACTAAA CTTCTTCGTA AAATTGTGGC CACAAGTCCA AATCCAATCA GAAAAATAAG 4140

AAGTTGCCCT AAAAATGTGA GCAAATTGAC TGTTAAGAGA GGACCTTTAG AAAAATCACT 4200

TAGTAGTTGA TAATAACGTA ATACCGCCAG GACAAGAATT GGCGTCAAAA GGGACTCTTT 4260

GATAGAACTG CGAGGTGCTC CCTTGAGAAT CTCTTTCATT ATTTTTTTAG GATTCTTACC 4320

TAGATAATCC TCTGCACTCA TGCCATCTCG TTCTGCTTCT GAGAAATCTA GCATCATCAA 4380

ATAGATCTGC TCTCTGAGAT AGTCTTCATC ATAGAGAAAT CCAGCAAGAT TAAAACTTTC 4440

CCACAACTCC TCAAAATACT TTTGATTCTC CTCAGAAAAC TCATGTAGCA AAGCGCTTGT 4500

TTCTTCGTAA TACTTCATTT TCTTCATGGT TTAACCCCCA TTCTTAATCC CTTCTACTTT 4560

TTGACTCAAA TCGTCCCATT GTTGCCAAAA GACTGAGACA CGCTCTTCTC CTTCTTTCAT 4620

TAATGAAAAA TACTTCCGAT CTGGACCATC TGGCGACGGG CGCATGTCGC CTCTTATCCA 4680

TTGATTTTTT TCTAACTTTT GCAACAAAGG ATAAATAGTT CCTGGAACGA TAGTATCAAA 4740

TCCAGCCTCT CGCAAAGTCT GAACCAACTC ATAACCATAC CGCTCTTTTT GACCAATCAT 4800

ATCCAAGACA CAACCTTCAA GAACACCTTT TAATAGCTGA GTTTCTTTCA TCACTTCTCC 4860

CTTCTAATCT ATTTTGTAAT ACCTACTAGT GACTTCACCT ATAGTATATC ACTTCTACAC 4920

TAGTTTGTAA AGCATAATAG TTAATACTCT TCGAAAATCT CTTCAAACCA CGTCAGCGTC 4980

GCCCTACCGT ATGTATGGTT ACTGACTTCG TCAGTTTCAT CTACAACCTC AAAAACATGT 5040

TTTGAGCTGA CTTCGTCAGT TTCATCTACA ACCTCAAAAC AGTGTTTTGA GCTGACTTCG 5100

TCAGTTTCAT CTACAACCTC AAAACAGTGT TTTGAGCTGA CTTCGTCAGT TTCATCTACA 5160

ACCTCAAAAA CATGTTTTGA GCTGACTTCG TCAGTTTCGT CTACAACCTC AAAACAGTGT 5220

TTTGAGCAAC CTGCGGCTAG CTTCCTAGTT TGCTCTTTGA TTTTCATTGA GTATAAATAA 5280

AAAAACAGAA CTAGCCTGAA CTAGTCCTGT CTACTTTTAC CCAATCACAC TTCCATTTGG 5340

TACAGCTGGA TCAACTGTGA GAAGGGTTAA TTTGCCATCA TGTTCAGCTG AGAGAATCAT 5400

ACCCTGGCTG ACATATTTTT TCATCATTTT ACGTGGTTTG AGGTTAGCAA CGATTTGAAC 5460

TTTCTTGCCG ACCAATTCTT GTTCATTTGG ATAGTATTTT GCAATTCCTG AAAGAATCTG 5520

ACGATCTTCT CCATCACCAG CATCCAAGCG GAATTGAAGC AACTTATCTG AACCTTCTAC 5580

TTTAGACACT TCTTTGACTT CTGCGACACG GATTTCAACC TTGTCAAAGT CTTCAAACTT 5640

GATTTCATCC TTGTTTAGTT TGAGCTCAAC TTCGTCCGGA TTCCATTCTT TTTCGACTGC 5700

TGGTTTATTG CCTTCCATTT GTTCCTTGAT ATAGGCGATT TCTTCTTCCA TATTTAGACG 5760

TGGAAAGATA GGTGTTCCTT TGGCAACTAC AGTCACATCT GCTGGGAAGT CAGCCAAACT 5820

CAAGTTTTCA AGACTAGAAA CTTCTTCCAA ACCAAGTTGA GTCAAAACTG CACGACTAGT 5880

TTCCATCATA AATGGTTCAA TCAAGTGAGC AACTACACGA ATGCTGGCTG CCAAGTGGCT 5940

CATGACACTT GCCAATTGGT CACGAAGAGC TTCATCCTTG GCCAAGACCC ATGGTGCGGT 6000

CTCATCGATG TATTTATTGG TACGAGAGAT CAGAGTCCAG ACTGCTTCAA GCGCACGTGG 6060

ATAGTCAACT GCTTCCATGT GTGTATGGAA GTCTGCGATT GATTGTwCTG CAACCTCAGC 6120

AAGAACATGA TCATATTCAG TCACACCTTC TACATAGGCA GGGATTTGTC CATCAAAGTA 6180

CTTATTAATC ATGGAAACCG TACGGTTAAG GAGGTTCCCA AGGTCATTAG CCAATTCATA 6240

GTTGATACGG CCGACATAGT CTTCAGGAGT AAAGGTTCCG TCTGAACCAA CTGGAAGGTT 6300

ACGCATGAGG TAGTAACGAA GTGGATCTAG TCCATAACGC TCTACCAACA TTTCAGGGTA 6360

AACGACATTC CCTTTTGACT TAGACATTTT TCCGTCTTTC ATGACAAACC AACCATGGGC 6420

AATCAAACGA TCAGGTAATT TAACATCCAA CATCATAAGA AGGATTGGCC AGTAGATAGA 6480

GTGGAAGCGA AGGATATCTT TTCCTACCAT ATGGAAGACT GTTCCATTCC AGAACTTGTC 6540

AAAGTTACCA TGTTCGTCTT GAGCGTAGCC AAGAGCTGTC GCATAGTTAA GAAGGGCATC 6600

AATCCAAACG TAGACAACGT GTTTTGGATT TGATGGGACA GGCACTCCCC ATGTAAAGGT 6660

TGTACGAGAT ACCGCCAAAT CTTCCAAGCC TGGCTCGATG AAGTTGCGTA GCATTTCATT 6720

AAGGCGACCA TCTGGCGTGA TAAATTCAGG ATGAGCTTTG AAAAATTCGA CCAAACGGTC 6780

TTGGTATTTG CTAAGGCGAA GGAAGTATGA TTCTTCAGAA AcCCATTCAA CCTCATGACC 6840

TGATGGAGCA ATACCACCAG TCACATTTCC AGCTTCATCA CGGAAAACTT CTGCCAGCTG 6900

GCTTTCTGTA AAGAATTCTT CGTCTGATAC TGAATACCAA CCAGAGTATT CACCCAAGTA 6960

GATATCATCT TGAGCAAGTA AGCGTTCAAA GACTTGTGCG ACAACTTTTT CATGGTAGTC 7020

ATCAGTTGTA CGGATAAATT TATCGTATGA GATATCTAGT AATTGCCAGA GTTCTTTAAC 7080

TCCAACCGCC ATTCCATCAA CATAGGCTTG AGGTGTAATA CCAGCTTCTT CCGCTTTCTG 7140

CTGGATTTTC TGACCATGTT CATCAAGACC TGTCAGATAA AATACATCGT AGCCCATCAG 7200

GCGTTTGTAA CGTGCTAGGA CATCACATGC GATAGTTGTG TAGGCAGAAC CGATATGAAG 7260

TTTCCCAGAT GGATAGTAAA TCGGCGTTGT AATATAAAAA TTTTTTTCAG ACATAATTTT 7320

TCCTTTCCAG GCAAATGAAA CCTGTTTTTC TAACACTTCA TTATATCACA TTTTTAATGA 7380

ATTTCAATAG GGAAATCCAT ACAAAAACAA GATAGACGAG TGTCCATCTT GTTGATCTCA 7440

TTCATAACGA AGGGCTTCAA TTGGATCAAG TTTCGATGCC TTGTTGGCTG GCAAGACTCC 7500

AAAAATCATA CCAACACTAG CCGAAACTGC AAGACTAAAT AGGGCGACTG GGATTGATAC 7560

TCCAACTTCT ATACCTTCTA TTAAACCTTG CAGTAACAAA CCTGCTAAGg CAGTTAAACC 7620

ACTTGCAATT GTCAAGCCAA TTAAGCCACC TAACAAGGTC AAAATCATGG ATTCAATCAA 7680

AAACTGAATT AAAATATTGG CACGTGTTGC ACCCAAAGCC TTACGAAGAC CAATCTCACG 7740

AGTGCGCTCT GTCACCGAAA CCAGCATGAT GTTCATGACA CCAGTTCCTC CAACAAAGAG 7800

AGAAATCCCT GCGATGGAAC TAATAATCGT CGTCATAAAA CTAAACGATT GTTGAATTTC 7860

TGCAAATACA ACGGACTCAT CTGCCACCTG GTATTCTCCC TGTTGTAAGC CTGCAAGCTC 7920

TGTCATTTTT CGTGCCAGTT CTGGACCCAG AGTTGGGGTT AAACTGGTAT CATTCACTCG 7980

AAAGACAATA TTAGCTATTT CATCTACATT AAAATTCGCA GCAAGGGAGA TATTGGTAGT 8040

AATAGGCAAG CCACCAAACC CATATATTTT TGATCTTTTA GCCTCCGGAC TAGTATAAAC 8100

CCCAATGACC CGGTAACTAA ATCCATTGAC TTCTACAACC TTGTTAATAG CCTCTTGAGG 8160

AGATTCAAAT AAACTAATGG ACAATTCCTC ATCTAGCAAA ATGACACTTG CAAACTCTTT 8220

GAAATCTTGC TCTCTCAGAC TACGACCTGC AATAATTTCA TTCTTAACAG CGTCCATGTA 8280

AGTTCTGTTT CCACCTGTCA AATTAGCATT CTCAACCTTT TTATCTTGAT AGGTCAAGAT 8340

GGCATTCGTT GAATTGGTTA CATAGTAACT ATCCACTCCC TTCAGTTTAG CTGCCTCTTG 8400

GACCCAGGAT TCTTGCGGTT TTGGCGGTTC AACAGGAACT TCCTCTTCCT TTCCAGAAAC 8460

CGTAAAAGCT GATTGTTTCT GAGTAAAAGA CCCGTCTTTA CTTTTTTTAG GAGAGAAAAA 8520

GACGCTAATA TTTTTCTGAG ATTTAGTCAT ATCTTTATTG ACTTGACGAG ATAGGGAATC 8580

ACCCAAAGCC ATAATCACAA CAACTGATGA AACACCGATA ATAATCCCAA TCATAGTAAG 8640

CAAAGAACGC ATCTTGTGAG CCATGATAGA TGAAAAGGCA AATTTCAGAT TCTGCATCTT 8700

AGTTTTCCTC CTTTCCTAAC TGAGCACTGT CAGACGAAAT GACCCCATCC CGAATGACAA 8760

TCTGACGTTT GGCATAGGCA GCAATCTCAG GCTCATGCGT TACCATGATA ATGGTTTTTC 8820

CTTCTTTATT CAAATCAACC AATAATTGCA TAATTTGGTT ACCTGTTTTG GTATCCAAGG 8880

CTCCTGTCGG TTCATCCGCT AGGATAATAG AAGGATTGTT TACCAAGGCA CGCGCAATGG 8940

CTACACGTTG CTTTTGACCA CCAGATAATT CTGAAGGTAA ATGGTGACTA CGTTCTGTCA 9000

ATTCAACCTT GTCTAAATAT TCCTCAGCCA ACTTGCGACG TTTTGAAGAC GAAACTCCTG 9060

CGTAAATCAA GGGCAATTCT ACATTTTGCA GAGCATTGAG CTTCGATAGA AGAAAGAACT 9120

GCTGAAAGAC AAAACCGATT TGTTGGTTAC GGACCTTAGC TAGTTGTTTT TCACCAAGCC 9180

CAGCCACTTC TTGACCTTCA AGATAATATT CTCCACTGGT TGGTGTATCC AACATGCCAA 9240

TCGTATTCAT CAGAGTGGAC TTACCAGACC CAGATGGTCC CATGATGGCT ACAAATTCAC 9300

CCTCATTCAC TTCTAGATTG ATATTTTTGA GAACCTGCAG TTCTTGGTCA CCATTACGGT 9360

AACTTCTGAA GATATTTTTT AGACTAATTA GTTGCTTCAT CAGCCTTCAC CTCTTTTCCT 9420

TCTTCCAAGG AAGATGTTGG ATTACTGATG ACCTTAGCAC CGTTCGTTAA ACCAGAAGTG 9480

ATTTCTTGAT TTTCTGCGTC AGCATTTCCC AATGAAACCT CAACTTTTTT AGCCTTTTGT 9540

TGTTCATCCA CAATCCAGAC ATAATTTTTA CTATCATCCA TTACTAGACT GCTAACAGGA 9600

ACAAGAATAG CCTTAGTTTT GCTTTTAACC TCAATGTTGA CAGAAAAACC TTGTTTCAAA 9660

TCACCAACCT CGCCTGTCAC ATCAATAGTA TAAGGGTATT TAGAACCTGT ATTATTCCCG 9720

GCTGCTGGAC TAGCTGCTTC ACCATTGTTT TTAGGATAGT CAGAAATATA GCTTAATTTC 9780

CCAGTCCATT TTTTATCAGG ATACACTTTA GAAGTAAAGC TTACTTCTTG ACCTACAGAA 9840

AGGTTGGCTA GATTGTACTC AGACAATTCT CCCTTGACTT GTAAATTTTC ATTGCTGACA 9900

ATATGAACCA TAACTTGACT CGCCCCTGTT GGAGATTTAG AAACATTGCT ATTGACTTCG 9960

ACCACAGTTC CCTCTAGGGT ACTGAGAACA GTTGTTGCAT CCAATTGACT TTGAGCCTTG 10020

CTTAATTGCG CCGCAGCATC TGCACGCGCA TCACGGGCAT CACCCAATTG AGCGTCAATA 10080

GAAGCAACAG AATTTCCAGC CACTGGAGTT GGGCTTTGCA CCGTTGCATC TTCTCCTCCT 10140

ACTGGCGCTG GTAACTGTGG AGCCGGAGCT GAAGCGGCTT CATTTCGTGC TTGATTGAGT 10200

TCATTGATAT GACGATCTGC CCTAGCTACT GCTCGACTAG CTGAATCATA GGCCGCCTGC 10260

GCTTCTGAAC TACTGTACTT GACTAAAGCC TGCCCTTCGC TGACCTTATC GCCCACAGAA 10320

ACAAGGATTT CATCTAAATC ACCCTTACTA GCATCAAAAT AAACATATTG TTCATTTTTT 10380

GCTGTTACTG TCCCTGACAA TAAAACAGAG GAGGCCACGC TTCCTTCCTT GGCAACAACA 10440

AGATGAGTAG GCTCATCTTT TAGAGCAGTC TGAGAAGGTT GTCTAAAGAG TAAAATCCCC 10500

CCAGCACCCA ATACAACTAC ACTCGCAGCA CCGATTGCTG CATACAGTTG CCACTTTTTA 10560

GCTTTACCAT TCTTTTTCTT CATAATGAAA CTCCTTTTCT TTTTTACAAT ACTTTGCTAT 10620

TATACCAAAT TTCCCTCCAG CAAACAATAC AGTTCAGGAT TAAACAATCG TTCGGAATTT 10680

TGCTTTTCGG 10690

(2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8195 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

GAGAAAGCGC CCACGTTTCC CCGAAGGGAG AAAGGCGGAC AGGTATCCGG TAAGCGGCCA 60

GGGTCGGAAC AGGAGAGCGC AACGAGGGAG CTTCCCAGGG GGAAACGCCT GGTATCTTTA 120

TAGTCCTGTC GGGTTTCGCC ACCTCTGACT TGAGCGTCGA TTTTTGTGAT GCTCGTCAGG 180

GGGGCGGAGC CTATGGAAAA ACGCCAGCAA CGCGGCCTTT TTACGGTTCC TGGCCTTTTG 240

CTGGCCTTTT GCTCACATGT TCTTTCCTGC GTTATCCCCT GATTCTGTGG ATAACCGTAT 300

TACCGCCTTT GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCGAGC GCAGCGAGTC 360

AGTGAGCGAG GAAGCGGAAG AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC 420

GATTCATTAA TGCAGCTGGC ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA 480

CGCAATTAAT GTGAGTTAGC TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC 540

GGCTCGTATG TTGTGTGGAA TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA 600

CaTGATTACG AATTCGAGCT CGGTACCCGG AAAATCCAGA AAATGCTTGA AAAAAATCCT 660

AGAAGATGGT ATAATACTAA ATTGTAAGGG TTATCACATA TAACTCAAAA AAAGAAAGAA 720

CAAAAGGAGA GTCAAACTAT GGCTTCTAAA GATTTCCACG TAGTGGCAGA AACAGGTATT 780

CACGCACGTC CAGCAACATT GTTGGTACAA ACTGCTAGCA AATTTGCTTC AGATATCACT 840

CTTGAGTACA AAGGTAAATC AGTTAACCTT AAATCAATTA TGGGTGTTAT GAGTCTTGGT 900

GTTGGCCAAG GTGCTGACGT AACTATCTCA GCTGAAGGTG CAGATGCAGA TGACGCTATC 960

GCTGCAATCT CAGAAACAAT GGAAAAAGAA GGATTGGCAT AAGGGAAATG ACAGAAATGC 1020

TTAAAGGAAT CGCAGCATCT GACGGTGTTG CAGTTGCAAA AGCATATCTA CTCGTTCAGC 1080

CGGATTTGTC ATTTGAGACT ATTACAGTCG AAGATACAAA CGCAGAAGAA GCTCGCCTTG 1140

ATGCCGCTCT ACAGGCATCA CAAGACGAGC TTTCTGTTAT TCGCGAGAAA GCAGTAGGTA 1200

CGCTCGGTGA AGAAGCAGCT CAAGTTTTTG ATGCTCACTT AATGGTTCTT GCTGACCCAG 1260

AAATGATCAG CCAAATCAAG GAAACTATCC GTGCGAAGAA AGTGAATGCA GAAGCAGGTC 1320

TGAAAGAAGT TACAGATATG TTTATCACTA TCTTTGAAGG CATGGAAGAC AACCCATACA 1380

TGCAAGAACG CGCAGcGGAT wTCCGCGACG TGACAAAACG TGTATTGGCA AACCTTCTTG 1440

GTAAAAAATT GCCAAACCCA GCTTCTATCA ATGAAGAAGT GATTGTGATT GCGCATGACT 1500

TGACTCCTTC AGATACAGCT CAATTGGACA AAAACTTTGT AAAAGCTTTT GTAACCAACA 1560

TTGGTGGACG TACAAGCCAC TCAGCTATCA TGGCACGTAC ACTTGAAATT GCTGCTGTAT 1620

TAGGTACAAA TAACATCACT GAAATCGTTA AAGACGGTGA CATCCTTGCT GTTAACGGGA 1680

TCACTGGAGA AGTGATTATC AACCCAACAG ATGAACAAGC GGCAGAATTT AAAGCAGCTG 1740

GTGAAGCCTA TGCGAAACAA AAAGCTGAAT GGGCACTTTT GAAAGATGCT CAAACAGTGA 1800

CTGCTGACGG TAAACACTTC GAGTTGGCTG CTAATATCGG TACTCCAAAA GACGTTGAAG 1860

GTGTTAACAA CAACGGTGCA GAAGCTGTTG GACTTTACCG TACAGAGTTC TTGTACATGG 1920

ATTCTCAAGA CTTCCCAACT GAAGATGAGC AGTATGAAGC ATACAAGGCT GTTCTTGAAG 1980

GAATGAACGG TAAACCTGTT GTCGTTCGTA CAATGGATAT CGGTGGAGAT AAGGAACTTC 2040

CTTACTTCGA TATGCCTCAC GAAATGAACC CATTCCTTGG ATTCCGTGCT CTTCGTATCT 2100

CTATCTCTGA GACTGGAGAT GCTATGTTCC GCACACAAAT CCGTGCTCTT CTTCGTGCGT 2160

CTGTTCACGG TCAATTGCGT ATCATGTTCC CAATGGTTGC GCTCTTGAAA GAATTCCGTG 2220

CAGCGAAAGC AGTCTTTGAT GAAGAAAAAG CAAACCTTCT TGCTGAAGGT GTTGCAGTTG 2280

CGGATAACAT CCAAGTTGGT ATCATGATCG AGATTCCTGC AGCGGCTATG CTTGCAGACC 2340

AATTTGCTAA AGAAGTTGAC TTCTTCTCAA TTGGTACAAA CGACTTGATC CAATATACAA 2400

TGGCAGCAGA CCGTATGAAC GAACAAGTTT CATACCTTTA CCAACCATAC AACCCATCAA 2460

TCCTACGCTT GATTAACAAT GTGATCAAAG CAGCTCACGC TGAAGGTAAA TGGGCTGGTA 2520

TGTGTGGTGA GATGGCTGGT GACCAACAAG CTGTTCCACT TCTTGTCGGA ATGGGCTTGG 2580

ATGAGTTCTC TATGTCAGCA ACATCTGTAC TTCGTACACG CAGCTTGATG AAGAAACTCG 2640

ACACAGCTAA GATGGAAGAG TACGCAAACC GTGCCCTTAC AGAATGCTCA ACAATGGAAG 2700

AAGTTCTTGA ACTTCAAAAA GAATACGTTA ATTTTGATTA ATCGAAAAGT CCCTGCAACT 2760

CAGTTACAGG GATTTTTTTG ATATTTTAAA AAGAATTTTC AAGAAAATCT TTCTTATAGA 2820

AAGTCCAACC TTGAAAAAGT AGTGGTCAGA ACAAAAAATA CTTAAATGGT TCATAAAATT 2880

CTTGACAAGT TGGATATTTA GGAGTAAACT ATTAACCAGT TAAGTAATAG AGAGGAGTTT 2940

CTGCAATTTA GAAATGAATT GCAACTAGAA ATATCAAATA GAAAGAGAGT TTCGATGAAA 3000

ATTAATAAGA AATACCTTGT TGGTTCTGCG GCACTTTGAT TTTAAGTGTT TGTTCTTACG 3060

AGTTGGGACT GTATCAAGCT AGAACGGTTA AGGAAAATAA TCGTGTTTCC TATATAGATG 3120

GAAAACAAGC GACGCAAAAA ACGGAGAATT TGACTCCTGA TGAGGTTAGC AAGCGTGAAG 3180

GAATCAATGC TGAGCAAATC GTCATCAAGA TAACAGACCA AGGCTATGTC ACTTCACATG 3240

GCGACCACTA TCATTATTAC AATGGTAAGG TTCCTTATGA CGCTATCATC AGTGAAGAAT 3300

TACTCATGAA AGATCCAAAC TATAAGCTAA AAGATGAGGA TATTGTTAAT GAGGTCAAGG 3360

GTGGATATGT TATCAAGGTA GATGGAAAAT ACTATGTTTA CCTTAAGGAT GCTGCCCACG 3420

CGGATAACGT CCGTACAAAA GAGGAAATCA ATCGACAAAA ACAAGAGCAT AGTCAACATC 3480

GTGAAGGTGG AACTCCAAGA AACGATGGTG CTGTTGCCTT GGCACGTTCG CAAGGACGCT 3540

ATACTACAGA TGATGGTTAT ATCTTTAATG CTTCTGATAT CATAGAGGAT ACTGGTGATG 3600

CTTATATCGT TCCTCATGGA GATCATTACC ATTACATTCC TAAGAATGAG TTATCAGCTA 3660

GCGAGTTGGC TGCTGCAGAA GCCTTCCTAT CTGGTCGAGG AAATCTGTCA AATTCAAGAA 3720

CCTATCGCCG ACAAAATAGC GATAACACTT CAAGAACAAA CTGGGTACCT TCTGTAAGCA 3780

ATCCAGGAAC TACAAATACT AACACAAGCA ACAACAGCAA CACTAACAGT CAAGCAAGTC 3840

AAAGTAATGA CATTGATAGT CTCTTGAAAC AGCTCTACAA ACTGCCTTTG AGTCAACGAC 3900

ATGTAGAATC TGATGGCCTT GTCTTTGATC CAGCACAAAT CACAAGTCGA ACAGCTAGAG 3960

GTGTTGCAGT GCCACACGGA GATCATTACC ACTTCATCCC TTACTCTCAA ATGTCTGAAT 4020

TGGAAGAACG AATCGCTCGT ATTATTCCCC TTCGTTATCG TTCAAACCAT TGGGTACCAG 4080

ATTCAAGGCC AGAACAACCA AGTCCACAAC CGACTCCGGA ACCTAGTCCA GGCCCGCAAC 4140

CTGCACCAAA TCTTAAAATA GACTCAAATT CTTCTTTGGT TAGTCAGCTG GTACGAAAAG 4200

TTGGGGAAGG ATATGTATTC GAAGAAAAGG GCATCTCTCG TTATGTCTTT GCGAAAGATT 4260

TACCATCTGA AACTGTTAAA AATCTTGAAA GCAAGTTATC AAAACAAGAG AGTGTTTCAC 4320

ACACTTTAAC TGCTAAAAAA GAAAATGTTG CTCCTCGTGA CCAAGAATTT TATGATAAAG 4380

CATATAATCT GTTAACTGAG GCTCATAAAG CCTTGTTTGA AAATAAGGGT CGTAATTCTG 4440

ATTTCCAAGC CTTAGACAAA TTATTAGAAC GCTTGAATGA TGAATCGACT AATAAAGAAA 4500

AATTGGTAGA TGATTTATTG GCATTCCTAG CACCAATTAC CCATCCAGAG CGACTTGGCA 4560

AACCAAATTC TCAAATTGAG TATACTGAAG ACGAAGTTCG TATTGCTCAA TTAGCTGATA 4620

AGTATACAAC GTCAGATGGT TACATTTTTG ATGAACATGA TATAATCAGT GATGAAGGAG 4680

ATGCATATGT AACGCCTCAT ATGGGCCATA GTCACTGGAT TGGAAAAGAT AGCCTTTCTG 4740

ATAAGGAAAA AGTTGCAGCT CAAGCCTATA CTAAAGAAAA AGGTATCCTA CCTCCATCTC 4800

CAGACGCAGA TGTTAAAGCA AATCCAACTG GAGATAGTGC AGCAGCTATT TACAATCGTG 4860

TGAAAGGGGA AAAACGAATT CCACTCGTTC GACTTCCATA TATGGTTGAG CATACAGTTG 4920

AGGTTAAAAA CGGTAATTTG ATTATTCCTC ATAAGGATCA TTACCATAAT ATTAAATTTG 4980

CTTGGTTTGA TGATCACACA TACAAAGCTC CAAATGGCTA TACCTTGGAA GATTTGTTTG 5040

CGACGATTAA GTACTACGTA GAACACCCTG ACGAACGTCC ACATTCTAAT GATGGATGGG 5100

GCAATGCCAG TGAGCATGTG TTAGGCAAGA AAGACCACAG TGAAGATCCA AATAAGAACT 5160

TCAAAGCGGA TGAAGAGCCA GTAGAGGAAA CACCTGCTGA GCCAGAAGTC CCTCAAGTAG 5220

AGACTGAAAA AGTAGAAGCC CAACTCAAAG AAGCAGAAGT TTTGCTTGCG AAAGTAACGG 5280

ATTCTAGTCT GAAAGCCAAT GCAACAGAAA CTCTAGCTGG TTTACGAAAT AATTTGACTC 5340

TTCAAATTAT GGATAACAAT AGTATCATGG CAGAAGCAGA AAAATTACTT GCGTTGTTAA 5400

AAGGAAGTAA TCCTTCATCT GTAAGTAAGG AAAAAATAAA CTAATGAAAA ATGAAAGTCT 5460

CGATAAAGAG GCTTTCATTT TTATTATGTA TATATGTAAA ATTCTTGACA AGCAATATTA 5520

AAAAGAGTAA ACTATTAACT AGTTAATTAA CCGGTTTATT ACTTTATAGT GAATCAAATA 5580

TACTTAAGAA AAGAGGAAAG AATGAAAATT AATAAAAAAT ATCTAGCAGG TTCAGTGGCA 5640

GTCCTTGCCC TAAGTGTTTG TTCCTATGAA CTTGGTCGTC ACCAAGCTGG TCAGGTTAAG 5700

AAAGAGTCTA ATCGAGTTkC TTATATAGAT GGTGATCAGG CTGGTCAAAA GGCAGAAAAC 5760

TTGACACCAG ATGAAGTCAG TAAGAGGGAG GGGATCAACG CCGAACAAAT CGTCATCAAG 5820

ATTACGGATC AAGGTTATGT GACCTCTCAT GGAGACCATT ATCATTACTA TAATGGCAAG 5880

GTCCCTTATG ATGCCATCAT CAGTGAAGAG CTCCTCATGA AAGATCCGAA TTATCAGTTG 5940

AAGGATTCAG ACATTGTCAA TGAAATCAAG GGTGGTTATG TTATCAAGGT AGATGGAAAA 6000

TACTATGTTT ACCTTAAGGA TGCAGCTCAT GCGGATAATA TTCGGACAAA AGAAGAGATT 6060

AAACGTCAGA AGCAGGAACA CAGTCATAAT CACGGGGGTG GTTCTAACGA TCAAGCAGTA 6120

GTTGCAGCCA GAGCCCAAGG ACGCTATACA ACGGATGATG GTTATATCTT CAATGCATCT 6180

GATATCATTG AGGACACGGG TGATGCTTAT ATCGTTCCTC ACGGCGACCA TTACCATTAC 6240

ATTCCTAAGA ATGAGTTATC AGCTAGCGAG TTAGCTGCTG CAGAAGCCTA TTGGAATGGG 6300

AAGCAGGGAT CTCGTCCTTC TTCAAGTTCT AGTTATAATG CAAATCCAGC TCAACCAAGA 6360

TTGTCAGAGA ACCACAATCT GACTGTCACT CCAACTTATC ATCAAAATCA AGGGGAAAAC 6420

ATTTCAAGCC TTTTACGTGA ATTGTATGCT AAACCCTTAT CAGAACGCCA TGTGGAATCT 6480

GATGGCCTTA TTTTCGACCC AGCGCAAATC ACAAGTCGAA CCGCCAGAGG TGTAGCTGTC 6540

CCTCATGGTA ACCATTACCA CTTTATCCCT TATGAACAAA TGTCTGAATT GGAAAAACGA 6600

ATTGCTCGTA TTATTCCCCT TCGTTATCGT TCAAACCATT GGGTACCAGA TTCAAGACCA 6660

GAACAACCAA GTCCACAATC GACTCCGGAA CCTAGTCCAA GTCCGCAACC TGCACCAAAT 6720

CCTCAACCAG CTCCAAGCAA TCCAATTGAT GAGAAATTGG TCAAAGAAGC TGTTCGAAAA 6780

GTAGGCGATG GTTATGTCTT TGAGGAGAAT GGAGTTTCTC GTTATATCCC AGCCAAGGAT 6840

CTTTCAGCAG AAACAGCAGC AGGCATTGAT AGCAAACTGG CCAAGCAGGA AAGTTTATCT 6900

CATAAGCTAG GAGCTAAGAA AACTGACCTC CCATCTAGTG ATCGAGAATT TTACAATAAG 6960

GCTTATGACT TACTAGCAAG AATTCACCAA GATTTACTTG ATAATAAAGG TCGACAAGTT 7020

GATTTTGAGG CTTTGGATAA CCTGTTGGAA CGACTCAAGG ATGTCyCAAG TGATAAAGTC 7080

AAGTTAGTGG ATGATATTCT TGCCTTCTTA GCTCCGATTC GTCATCCAGA ACGTTTAGGA 7140

AAACCAAAATG CGCAAATTAC CTACACTGAT GATGAGATTC AAGTAGCCAA GTTGGCAGGC 7200

AAGTACACAA CAGAAGACGG TTATATCTTT GATCCTCGTG ATATAACCAG TGATGAGGGG 7260

GATGCCTATG TAACTCCACA TATGACCCAT AGCCACTGGA TTAAAAAAAGA TAGTTTGTCT 7320

GAAGCTGAGA GAGCGGCAGC CCAGGCTTAT GCTAAAGAGA AAGGTTTGAC CCCTCCTTCG 7380

ACAGACCATC AGGATTCAGG AAATACTGAG GCAAAAGGAG CAGAAGCTAT CTACAACCGC 7440

GTGAAAGCAG CTAAGAAGGT GCCACTTGAT CGTATGCCTT ACAATCTTCA ATATACTGTA 7500

GAAGTCAAAA ACGGTAGTTT AATCATACCT CATTATGACC ATTACCATAA CATCAAATTT 7560

GAGTGGTTTG ACGAAGGCCT TTATGAGGCA CCTAAGGGGT ATACTCTTGA GGATCTTTTG 7620

GCGACTGTCA AGTACTATGT CGAACATCCA AACGAACGTC CGCATTCAGA TAATGGTTTT 7680

GGTAACGCTA GCGACCATGT TCGTAAAAAT AAGGTAGACC AAGACAGTAA ACCTGATGAA 7740

GATAAGGAAC ATGATGAAGT AAGTGAGCCA ACTCACCCTG AATCTGATGA AAAAGAGAAT 7800

CACGCTGGTT TAAATCCTTC AGCAGATAAT CTTTATAAAC CAAGCACTGA TACGGAAGAG 7860

ACAGAGGAAG AAGCTGAAGA TACCACAGAT GAGGCTGAAA TTCCTCAAGT AGAGAATTCT 7920

GTTATTAACG CTAAGATAGC AGATGCGGAG GCCTTGCTAG AAAAAGTAAC AGATCCTAGT 7980

ATTAGACAAA ATGCTATGGA GACATTGACT GGTCTAAAAA GTAGTCTTCT TCTCGGAACG 8040

AAAGATAATA ACACTATTTC AGCAGAAGTA GATAGTCTCT TGGCTTTGTT AAAAGAAAGT 8100

CAACCGGCTC CTATACAGTA GTAAAATGAA TGGAGCATAT TTTATGGAGA AGTAACCTTT 8160

CGTGTTACTT CTCTTTTTTA GAAAAACGTA ACAGA 8195

(2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2004 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

TTTACTAAAA GGAAAAAAGA ACTGATTTCT CAGTCCTTCA TTAATCTTAT TCCACACTAA 60

ATAGGTATGG GTAAACAGGT TGTTGACCTT GGTGAATCTC GACTTCAACG TCTTCGAATT 120

CTTCTACGAT TTCTTGAGCG ATTTCATTGG CAAGTTCTTC GCTTCCGTCT TCACCTACAT 180

AGAAGGTTAC GATTTCACTG TCTTCATCCA ACATATGTTT CAAGGTTTCA GTCAATGTTT 240

GGTGCATATC AGGGTTTGAC ACAAGAATTT TTCCATCCAC CATACCTAAA TTATCGTTTT 300

CATGGATTTC TAAGCCATCG ATCGTTGTAT CACGCACGGC TGTTGTGACG CTTCCGCTAA 360

CGACATCGCT AAGAGCAGCT GTCATACGCT CTTGGTTTTC TTCAATGGAC TTGCTTGGAT 420

CAAAGGCAAG AAGACTTGTC ATACCTTGAG GAAGAGTGCG AGCCTCTACC ACTACCGCTG 480

GTTGCTCCAA AACTTCTGCC GCAGATTGAG CTGCCATGAA GATGTTCTTG TTGTTTGGCA 540

AGAAGATGAT GTTACGGGCA TTAACCTGTT CAACAGCCTT GATAAAGTCT TCTGTTGAAG 600

GGTTCATGGT TTGACCGCCT TCGATAACAT AATCCACGCC TTGAGAACAG AAGATATCTG 660

CTAGACCTTT ACCAGCCACC ACAGCAATCA AAGCATACTC TTTTTCTTCA GCCGACTTGA 720

TAACTTGAGT AGCTTCTTTC TCAACCTGTG CTTCGTGTTG GTTACGCATA TTGTCAACTT 780

TTACCTTGAC CAAGCTACCA TATTTGAGAC CTTCTTGCAT AACAAGTCCT GGATCTTCTG 840

TATGAACATG GACTTTGACA ATTTCATCAT CGTTAACAAC AAGGAGAGAA TCTCCAAGCT 900

CATCCAAGTA GTTACGGAAT TCATCGTAGT CAAAATCTTT AGCATAGGTT GGACCTTGCT 960

TAAGAGCTAC CATGATTTCA GTACAGTAAC CAAACGTGAT GTCCTCAGTC GCTACGTGAC 1020

CAGCTACAGA CTTATGATGC TCTACATTGA TCATCTCACT CATGTTGGCA GGAGTCGCTA 1080

CAAAGTCCTC AGATGCAATA TATTCGCCAG TAAGGGCTGA AAGGAAACCT TCGTAGATGA 1140

AGACCAATCC TTGACCACCT GAGTCCACAA CGCCAACTTC TTTCAATACT GGAAGCATGT 1200

CTGGTGTTTT AGCTAGAGCT GTTTTAGCAC CTTCCAAGGC TGCGCGCATG ACTTCAACAG 1260

CGTCATCTGT TTGCTCAGCT TTTTTCTTAG CACCGATAGC AGCTCCACGA GAAACTGTTA 1320

AAATCGTTCC TTCAACAGGT TTCATCACTG CCTTATAGGC AACTTCCACA CCTGATTGGA 1380

AGGCCAGAGC CAAGTCTTGA CCTGTTAACT CGTCTTTATC CTTGATAGCT TGGGAAAATC 1440

CACGGAAAAG CTGAGACGTA ATCACTCCTG AGTTCCCACG CGCACCCATC AAAAGCCCTT 1500

TGGCAAGAAT GCTCGCTACT TCTCCAACTG TAGAAGCTGG CTTGTCTGCA ACTTCTTTAG 1560

CACCATTTTC AATGGTCATT CCCATATTTG TCCCAGTATC TCCATCTGGA ACTGGAAAGA 1620

CGTTTAATGA ATTGACATAT TCAGCTTGCT TATTCAAGCG AGTTGATGCA GCCTGCACCA 1680

TTTCTTGAAA TAAGCTAGTA GTAATTTTTG ACACGGTTAT TCTCCTACAA CTTTGATATT 1740

TTGAATGTAG ACATTTACAG TCTGAGCAGT AATTCCAAGC TGGTTTTCCA AGCTAAAGGC 1800

AACACGCTCT TGAATGTTTT TTGACACTTC ACTAATCTTT GTTCCGTAGC TTAACACGGT 1860

ATATACATCA ACTGCAATAC TGCCATCTTC GGCTGCCTTT ACGACGACAC CTTTAGAATA 1920

ATTTTCCTTA CCTAGCAGGG CTTGGAAATT ATCTTTGAGG GCATTTTTAC TAGCCATACC 1980

GACCACACCA GAAATCTCAG TTGC 2004

(2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11915 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

CCGGGTTGGG CTGTTCGCCC ATTAAAGCGG CACCACAGCT GGGTTCAGAA CGTCGTGAGA 60

CAGTTCGGTC CCTATCCGTC GCGGGCGTAG GAAATTTGAG AGGATCTGCT CCTAGTACGA 120

GAGGACCAGA GTGGACTTAC CGCTGGTGTA CCAGTTGTCT TGCCAAAGGC ATCGCTGGGT 180

AGCTATGTAG GGAAGGGATA AACGCTGAAA GCATCTAAGT GTGAAACCCA CCTCAAGATG 240

AGATTTCCCA TGATTATATA TCAGTAAGAG CCCTGAGAGA TGATCAGGTA GATAGGTTAG 300

AAGTGGAAGT GTGGCGACAC ATGTAGCGGA CTAATACTAA TAGCTCGAGG ACTTATCCAA 360

AGTAACTGAG AATATGAAAG CGAACGGTTT TCTTAAATTG AATAGATATT CAATTTTGAG 420

TAGGTATTAC TCAGAGTTAA GTGACGATAG CCTAGGAGAT ACACCTGTAC CCATGCCGAA 480

CACAGAAGTT AAGCCCTAGA ACGCCGGAAG TAGTTGGGGG TTGCCCCCTG TGAGATAGGG 540

AAGTCGCTTA GCTCTAGGGA GTTTAGCTCA GCTGGGAGAG CATCTGCCTT ACAAGCAGAG 600

GGTCAGCGGT TCGATCCCGT TAACTCCCAT TTTAGCGGGT GTAGTTTAGT GGTAAAACTA 660

CAGCCTTCCA AGCTGTTGTC GCGAGTTCGA TTCTCGTCAC CCGCTTTGAA CTTTGTTCTT 720

TGTACCAAGT TTTTGACTTG GGCGCGTAGC TCAGGTGGTT AGAGCGCACG CCTGATAAGC 780

GTGAGGTCGG TGGTTCGAGT CCACTCGTGC CCATAGTGTT TAGTCCATTA CTAGGGGATT 840

GGAATATTAT CTGTTCACTA AGAGGACACG GGCTTGTTCC CGTATAAACT ATTTTGGAGG 900

ATTACCCAAG TCCGGCTGAA GGGAACGGTC TTGAAAACCG TCAGGCGTGT AAAAGCGTGC 960

GTGGGTTCGA ATCCCACATC CTCCTTTTAT ATTAACGCGG GATGGAGCAG CTCGGTAGCT 1020

CGTCGGGCTC ATAACCCGAA GGTCGTAGGT TCAAATCCTG CTCCCGCAAT AAGGCTCGGT 1080

AGCTCAGTTG GTAGAGCAAT GGATTGAAGC TCCATGTGTC GGCGGTTCGA TTCCGTCTCG 1140

CGCCATTTAT ATATTTTGGA AGGGTAGCGA AGAGGCTAAA CGCGGCGGAC TGTAAATCCG 1200

CTCCTTCGGG TTCGGGGGTT CGAATCCCTC CCCTTCCATT TTACGGGCAT AGTTTAAAGG 1260

TAGAACTAAG GTCTCCAAAA CCTTCAGTGT GGGTTCAATT CCTACTGCCC GTGTTAATAG 1320

AATTATGGCG GGTGTGGTGA AGTGGTTAAC ACACCAGATT GTGGCTCTGG CATGCGTGGG 1380

TTCGATCCCC ATCACTCGCC TATTTTATAT TGGGGTATAG CCAAGCGGTA AGGCAAGGGA 1440

CTTTGACTCC CTCATGCGTT GGTTCGAATC CAGCTACCCC AGTTACTATT TGCCGGCGTG 1500

GCGGAATTGG CAGACGCGCT GGACTCAAAA TCCAGTGTCC GCAAGGACGT GCCGGTTCGA 1560

CCCCGGCCGC CGGTATAGTA TAGTGTTAGG AACGTTGTTA TTCTTCGTTC CTTTTTTATA 1620

TTATTTTTGG TATAATTATA GTTATTCAAA TTTTATTTAG ATTAAGAAAG TGTAGGGGAG 1680

TATGTCTTGT TCTATCGATT TATTAAAACA TCGGTATTTG AAAAATATTA AAGAAAATCC 1740

TGAATTGTTT GTCGGAATTG AGTTGGAGTA TCCTGTTGCA AGTTTAGAAG GGGATGCTAC 1800

AGATGTTGAA GTTATGAAGG ATCTATTTCA TTATTTAGTT TCTACTTTGG ATCTCACCGT 1860

AGCAAAGGTA GATGATTTTG GCAATCTGAT CCAGTTAGTA GATCCGATAA GTCAGGATGC 1920

TATTTTATTT GAAGTTTCCT ATACAACGAT TGAGTTTGCA TTTGGTAAGG CTGAAACGAT 1980

TCAAGAGGTC GAAAATCGTT TCAATAATTA TATGAATGTA ATTCAGAGAA AGTTAGCTGA 2040

ATCAAATCAT GCTATTGTTG GCTGTGGTAT CCATCCCAAC TGGGATAAAA ATGAGAATTG 2100

TCCAGTGGCT TATCCACGCT ATCAGATGTT GATGGATTAT TTGAATTTGA GTAGAAATAT 2160

TATTAAATCA GATTTACATC ATTTCCCTGA ATATGGTACT TTTATCTGTG GGAGCCAGGT 2220

TCAGCTGGAT ATTTCAAAAA CCAACTACTT ACGGGTGATT AATGCTTTTA CTCAAATTGA 2280

AGCGGCTAAG GCTTATTTAT TTGCAAACTC TGAATTTTCG GGTGCGGATT GGGATACGAA 2340

AATTTCAAGG GATATTTTCT GGGAAGAATC TATGCATGGT ATCTATCCAG AGAATGTTGG 2400

GGTCAATGCT AGACTCCTTA ATGATGAAAC TGATTTTTTT GACTATCTAA ATCATTCTGC 2460

GATTTTTACT GCGGAACGTG ATGGGCAGAC CTATTATTTT TATCCTATTC AGGCTGGGGA 2520

CTATTTGGCT ACGTCCGAAA TCCAAGCATT TGCTCTGAAT GGGGATGAGG TTATTATTTA 2580

CCCCCAAGAG AAGGATTTTG AAACTCATCG TAGTTACCAG TACCAAGATT TAACGACTCG 2640

AGGAACAGTT GAGTTTCGTA GTGTGTGTAC ACAGCCACTT GATAGGACTT TTGCTTCTGC 2700

AGCTTTTCAC TTGGGATTAT TGGTTAATTT AGACAAGTTA GAAGCTTACT TAGAAACAGC 2760

ACCTTTCTTT AAAGTATTTG GTTATGATTA CAAGTCTTTA AGGAGACAAT TTTCTAAGAA 2820

AAATCTTACA GATGAGGAAG AAACTACGAT TATTGAATTT TCCAAAGACT TACTCCTACT 2880

AGCTGAGGAG GGACTAGTGG TGAGAAATAA GGAAGAAATG ACCTATTTAC AGCCTTTGAG 2940

AGAAGAATTG AGCCTATAAT TTCTCTTATA AAGGGAGAAT TTTCTGAAAA ATCATGATAT 3000

AATGGACGAG ACTATAGATA AAGGATAGAG AGTAATGACA TTAGTTTATC AATCAACGCG 3060

TGATGCCAAC AATACAGTAA CTGCCAGCCA AGCAATTTTG CAAGGTTTGG CGACGGACGG 3120

CGGTTTGTTT ACACCGGATA CTTATCCAAA GGTAGATTTG AACTTTGACA AATTGAAAGA 3180

TGCTTCTTAC CAGGAAGTTG CTAAGCTAGT TTTGTCAGCA TTTTTAGATG ACTTTACAGT 3240

TGAGGAGTTG GACTACTGTA TCAACAATGC CTACGATAGC AAATTTGATA CTCCAGCTAT 3300

TGCACCATTA GTGAAATTAG ATGGGCAATA CAATTTGGAA CTTTTCCATG GTTCAACGAT 3360

TGCCTTTAAG GATATGGCCT TGTCTATTTT GCCATACTTT ATGACGACTG CTGCTAAGAA 3420

ACATGGTTTG GAGAACAAGA TTGTTATCTT GACAGCGACA TCTGGTGACA CGGGGAAAGC 3480

TGCTATGGCG GGGTTTGCGA ATGTGCCTGG TACTGAGATT ATCGTCTTTT ATCCAAAGGA 3540

TGGTGTCAGC AAGATTCAAG AGTTACAAAT GACCACTCAG ACTGGCGACA ATACTCATGT 3600

TATTGCTATT GATGGTAACT TTGACGATGC GCAAACAAAT GTGAAGCACA TGTTTAACGA 3660

CGTGGCTCTT CGTGAAAAAT TGACTACCAA CAAGTTGCAA TTTTCATCAG CTAACTCTAT 3720

GAACATTGGT CGTCTGGTGC CACAAATTGT TTATTATGTT TATGCTTACG CTCAATTGGT 3780

TAAGACTGGT GAAATTGTAG CTGGTGAAAA GGTTAACTTC ACAGTACCAA CAGGAAACTT 3840

TGGAAATATC TTGGCTGCCT TTTATGCCAA ACAAATTGGT TTGCCAGTTG GTAAATTAAT 3900

CTGTGCTTCA AATGACAACA ATGTTTTGAC AGACTTCTTT AAAACACGTG TCTATGACAA 3960

AAAACGTGAG TTTAAGGTAA CAACCAGCCC ATCTATGGAT ATCTTGGTAT CTTCAAACTT 4020

GGAGCGCTTG ATTTTCCATC TTTTGGGAAA TAATGCTGAA AAGACAACTG AACTTATGAA 4080

TGCCTTGAAC ACGCAAGGAC AATATAAGTT GACAGACTTT GATGCAGAGA TTTTGGACCT 4140

CTTTGCAGCT GAATATGCGA CTGAGGAAGA AACGGCAGCA GAGATCAAGC GTGTTTGTGA 4200

GTTAGATTCT TATATCGAGG ACCCTCATAC AGCTGTTGCT TCAGCAGTTT ATAAAAAATA 4260

CCAATCGGCC ACTGGAGATG TAACTAAGAC AGTGATTGCT TCAACAGCTA GTCCATACAA 4320

GTTCCCAGTA GTTGCAGTAG AAGCTGTAAC TGGAAAAGCA GGTTTAACAG ACTTTGAAGC 4380

CTTGGCTCAA TTACATGAAA TCTCAGGCGT TGCAGTGCCA CCAGCAGTTG ATGGGCTTGA 4440

AATAGCTCCA ATTCGTCACA AGACAACAGT GGCAGCTGCT GACATGCAAG CAGCGGTTGA 4500

GGCTTATTTA GGACTTTAAG ACAGAGGGAG CAAACTCGGT TGGGAAACCA ACTGAGTTTC 4560

TTTTCATCAG GAGGAGAGAT TGTTTAAGAA AAATAAAGAC ATTCTTAATA TTGCATTGCC 4620

AGCTATGGGT GAAAACTTTT TGCAGATGCT AATGGGAATG GTGGACAGTT ATTTGGTTGC 4680

TCATTTAGGA TTGATAGCTA TTTCAGGGGT TTCAGTAGCT GGTAATATTA TCACCATTTA 4740

TCAGGCGATT TTCATCGCTC TGGGAGCTGC TATTTCCAGT GTTATTTCAA AAAGCATAGG 4800

GCAGAAAGAC CAGTCGAAGT TGGCCTATCA TGTGACTGAG GCGTTGAAGA TTACCTTACT 4860

ATTAAGTTTC CTTTTAGGAT TTTTGTCCAT CTTCGCTGGG AAAGAGATGA TAGGACTTAT 4920

GGGGACGGAG AGGGATGTAG CTGAGAGTGG TGGACTGTAT CTATCTTTGG TAGGCGGATC 4980

GATTGTTCTC TTAGGTTTAA TGACTAGTCT AGGAGCCTTG ATTCGTGCAA CGCATAATCC 5040

ACGTCTGCCT CTCTATGTTA GTTTTTTATC CAATGCCTTG AATATTCTTT TTTCAAGTCT 5100

AGCTATTTTT GTTCTGGATA TGGGGATAGC TGGTGTTGCT TGGGGGACAA TTGTGTCTCG 5160

TTTGGTTGGT CTTGTGATTT TGTGGTCACA ATTAAAACTG CCTTATGGGA AGCCAACTTT 5220

TGGTTTAGAT AAGGAACTGT TGACCTTGGC TTTACCAGCA GCTGGAGAGC GACTTATGAT 5280

GAGGGCTGGA GATGTAGTGA TCATTGCCTT GGTCGTTTCT TTTGGGACGG AGGCAGTTGC 5340

TGGGAATGCA ATCGGAGAAG TCTTGACCCA GTTTAACTAT ATGCCTGCCT TTGGCGTCGC 5400

TACGGCAACG GTCATGCTGT TGGCCCGAGC AGTTGGAGAG GATGATTGGA AAAGAGTTGC 5460

TAGTTTGAGT AAACAAACCT TTTGGCTTTC TCTGTTCCTC ATGTTGCCCC TGTCCTTTAG 5520

TATATATGTC TTGGGTGTAC CATTAACTCA TCTCTATACG ACTGATTCTC TAGCGGTGGA 5580

GGCTAGTGTT CTAGTGACAC TGTTTTCACT ACTTGGGACC CCTATGACGA CAGGAACAGT 5640

CATCTATACG GCAGTCTGGC AGGGATTAGG AAATGCACGC CTCCCTTTTT ATGCGACAAG 5700

TATAGGAATG TGGTGTATCC GCATTGGGAC AGGATATCTG ATGGGGATTG TGCTTGGTTG 5760

GGGCTTGCCT GGTATTTGGG CAGGGTCTCT CTTGGATAAT GGTTTTCGCT GGTTATTTCT 5820

ACGCTATCGT TACCAGCGCT ATATGAGCTT GAAAGGATAG GAAATGCAAA AAACAGCTTT 5880

TATTTGGGAT TTAGACGGGA CTTTATTGGA CTCTTACGAA GCGATTTTAT CAGGGATTGA 5940

GGAGACTTTT GCTCAGTTTT CTATTCCTTA TGATAAGGAG AAGGTGAGAG AGTTTATCTT 6000

CAAGTATTCG GTGCAAGATT TGCTTGTGCG GGTGGCAGAA GATAGAAATC TGGATGTTGA 6060

GGTGCTAAAT CAGGTGCGTG CCCAGAGTCT GGCTGAGAAG AATGCTCAGG TAGTTTTGAT 6120

GCCAGGTGCG CGTGAGGTGC TAGCTTGGGC AGACGAATCA GGAATTCAGC AGTTTATATA 6180

TACTCATAAG GGGAACAACG CTTTTACCAT TCTCAAGGAC TTGGGGGTGG AATCCTATTT 6240

TACAGAGATT TTAACCAGTC AGAGTGGCTT TGTGCGGAAG CCAAGTCCAG AAGCGGCTAC 6300

CTATCTGCTA GATAAGTATC AGTTGAATTC TGATAATACT TATTATATAG GGGATCGGAC 6360

TCTGGATGTG GAATTTGCCC AGAATAGTGG GATTCAAAGC ATCAACTTTT TAGAGTCTAC 6420

TTATGAAGGG AATCACAGGA TTCAAGCGTT AGCAGATATT TCCCGTATTT TTGAGACTAA 6480

GTGATAAAAA GATTGTGTCA GTTTTGTGAC AGAGACCTAA CAAACTATTT CAAGTAACCT 6540

AGTTTGTTAC AAGGAATAGA CAGTTCTGTT AAATAGGCCC GAGAGGGCTT TTTTTCTACA 6600

TTTTTTGTGT TATGATAGAC AGGTACTCAT TTGAAAGGAA TTTGAAAGAA TGAAGAAAAG 6660

AATGTTATTA GCGTCAACAG TAGCCTTGTC ATTTGCCCCA GTATTGGCAA CTCAAGCAGA 6720

AGAAGTTCTT TGGACTGCAC GTAGTGTTGA GCAAATCCAA AACGATTTGA CTAAAACGGA 6780

CAACAAAACA AGTTATACCG TACAGTATGG TGATACTTTG AGCACCATTG CAGAAGCCTT 6840

GGGTGTAGAT GTCACAGTGC TTGCGAATCT GAACAAAATC ACTAATATGG ACTTGATTTT 6900

CCCAGAAACT GTTTTGACAA CGACTGTCAA TGAAGCAGAA GAAGTAACAG AAGTTGAAAT 6960

CCAAACACCT CAAGCAGACT CTAGTGAAGA AGTGACAACT GCGACAGCAG ATTTGACCAC 7020

TAATCAAGTG ACCGTTGATG ATCAAACTGT TCAGGTTGCA GACCTTTCTC AACCAATTGC 7080

AGAAGTTACA AAGACAGTGA TTGCTTCTGA AGAAGTGGCA CCATCTACGG GCACTTCTGT 7140

CCCAGAGGAG CAAACGACCG AAACAACTCG CCCAGTTGAA GAAGCAACTC CTCAGGAAAC 7200

GACTCCAGCT GAGAAGCAGG AAACACAAGC AAGCCCTCAA GCTGCATCAG CAGTGGAAGT 7260

AACTACAACA AGTTCAGAAG CAAAAGAAGT AGCATCATCA AATGGAGCTA CAGCAGCAGT 7320

TTCTACTTAT CAACCAGAAG AGACGAAAAT AATTTCAACA ACTTACGAGG CTCCAGCTGC 7380

GCCCGATTAT GCTGGACTTG CAGTAGCAAA ATCTGAAAAT GCAGGTCTTC AACCACAAAC 7440

AGCTGCCTTT AAAGAAGAAA TTGCTAACTT GTTTGGCATT ACATCCTTTA GTGGTTATCG 7500

TCCAGGAGAC AGTGGAGATC ACGGAAAAGG TTTGGCTATC GACTTTATGG TACCAGAACG 7560

TTCAGAATTA GGGGATAAGA TTGCGGAATA TGCTATTCAA AATATGGCCA GCCGTGGCAT 7620

TAGTTACATC ATCTGGAAAC AACGTTTCTA TGCTCCATTC GATAGCAAAT ATGGGCCAGC 7680

TAACACTTGG AACCCAATGC CAGACCGTGG TAGTGTGACA GAAAATCACT ATGATCACGT 7740

TCACGTTTCA ATGAATGGAT AAACCCGACT TGATAACATC ATTTTGACGA ATGAGATCTA 7800

GCTTTCGTGA TGGAAAGCGA TTCTCGTTCG TTTTTTCTTT GTCATACTCT TCGAAAATCT 7860

CTTCAAACCA CGTCAGTTTT ATCTGAAACT TCAAAGCTGT GCTTTGAGCA ACCTGCGACT 7920

AGCTTCCTAG TTTGCTTTTT GATTTTCATT GAGTATCAAT TTGAATGGAA AATGGAAAGT 7980

TATCATCTTG TAATGAGTTA AGCAACATTC TTGCAATCTA TTTTACTTTA TATCACAATT 8040

AATTAGTCAA ATATTGATAA ATCAATAAAA AGAGAGGGGA AGAAATGCTA GAGATTCAAG 8100

ATTTACTGTA TCAACTCCGC TTGTCTGAGC AAGCGAGTAC GCAATTGTTT GAAAAAAGGC 8160

TTGGGATTAG TTTGACACGG TATCAGATTT TACTGTTTTT GCTGGAGCAT TCTCCTTGTA 8220

ACCAAATGGC GGTTCAGGAG CGTTTGAAAA TTGATCAGGC TGCTTTGACA CGGCATTTCA 8280

AAATTTTGGA AACGGAAGGT TTGGTGGAGC GTCATCGTAA TCCTGAAAAT CAGCGGGAAG 8340

TGTTGGTAGA GGCTGCGAAG TATGCCAAGG AGCAGTTAGT GGTGAATCCC CCTCTGCAAC 8400

ATATCAGGGT TAAGGAAGAG ATAGAAAGTA TCTTAACAGA GTTTGAGAGA ACAGAACTCA 8460

GCCGTTTATT AAATAAATTG GTTTTGGGTA TTGAAAATAT AGAAATTTAA GGAGAAATAG 8520

ATGTCAATTA TTTTAACAAC GATCGTTGCT TTGGAGCATT TTTACATTTT TTATTTGGAA 8580

AGTATTGCCA CGCAATCAGA TGCGACTAGT CGTGTATTTA ATATGGAAAA GGAAGAATTG 8640

GCTCATCCGT CAGTAAGTTC ATTGTTCAAA AATCAAGGAA TTTATAAGGC TCTGCTAGGA 8700

GTCTTTCTCT TGTATGTCAT TTATTTCTCA CAGAATTTAG AAATTGTGAC TATTTTTGTC 8760

TTATTTGTGA TTGGTGCTGC GACTTACGGC TCTTTAACAG CGGATAAAAA AATTATTTTG 8820

AAACAAGGTG GATCAGCTAT TTTGGCCTTG ATTAGTATTT TACTCTTTAA ATACACTTGA 8880

AGGTCGATTC TAATCTCGCT AATCCTTTTT AATCCAGAAT AAGGGAAATA TGTTATACTT 8940

GTTTTTAAGA AAAAAGTCTC ATTGAATTGG TTTTGAGGAG TTAGAAATGA AAGTATTAGT 9000

GACAGGTTTT GAGCCCTTTG GAGGGGAAAA GGGCAATCCA GCTTTGGAGG CCATTAAAGG 9060

TTTACCAGCT GAAATCCATG GTGCTGAGGT CCGTTGGCTA GAGGTGCCGA CAGTTTTTCA 9120

CAAATCTGCT CAAGTATTGG AAGAAGAGAT GAATCGTTAT CAACCTGACT TTGTCCTTTG 9180

TATTGGGCAA GCTGGTGGAA GAACTAGTTT GACACCTGAA CGAGTGACCA TTAATCAAGA 9240

CGATGCATGC ATTTCTGATA ACGAAGATAA TCAACCGATT GACCGTCCCA TTCGCCCAGA 9300

TGGTGCTTCG GCCTACTTTA GTAGTTTGCC GATTAAAGCG ATGGTTCAAG CTATAAAAAA 9360

AGAGGGCTTA CCGGCCTCTG TTTCCAATAC GGCAGGGACT TTTGTCTGCA GCCATTTGAT 9420

GTATCAGGCT CTCTATTTGG TAGAAAAGAA ATCTCCATAT GTTAAGGCAG GTTTTATGCA 9480

TATTCCTTAT ATGATGGAAC AGGTGGTGAA CAGACCGACT ACTCCAGCTA TGAGTTTAGT 9540

GGATATTCGG CGAGGGATAG AAGCAGCAAT CGGCGCTATA ATAGAACATG GAGATCAGGA 9600

ACTCAAGTTG GTAGGCGGAG AAACTCATTG ATAGAAAAAA GCTTGAGGGG AAAAACCTTC 9660

AAGCTTTTGG ACGTTTTCGG GCCAATACTG CTCGGTAAAA CATAATTTTA GTGCATTGGA 9720

TATAAGGTAG GAGTGAAAAA CTAGCAATGC CAAAGGTAAT CCAATTGAGG AAGTACCAAG 9780

GAAGAAGCTG TAAATCTAGG ACAAAGTGCT GGAACTTGTA GCCCTTCATA AAGGAACGGC 9840

TAGTTTTTAG GATTCGTCTT GGTGGGACCT GTCCTAGGTC TAGACTATAA CAGAGAAGAA 9900

ATTCCACCTG TGAATAGGCA TAATACTGTG GAATATAGAG GATATTTCCT ACAATGATCA 9960

AGATGAGACT TGCAAGAAAG TAGAGTCCAA AGACCATGAG GAAACGCTCG GTTTCAACTG 10020

ATGAGAGATC TAGATTTGGA AACTCAGGAT GTAGGGTGAC GAATTTTTTG GCTAAAAAGC 10080

TACTATAAAA GAGGAGGTAA ATCCCAAGTA AATTAGGGAT ACTCCATAAA AAGAGATAGA 10140

AACGTTTGAG AAGTAGGGTC AAAAAGGTTT GAGAAAAGCG CTCCTCATCA AAGAGAGCTA 10200

GGCTGTTTTT TACAGATGGC TCCGTTTTAG AATCTTTCAT GAGTGTCAGT GTTGCATAGA 10260

CGGAACTGGT CAAAAGAATA GTCCCGATAA AGGAGACTAG TAGAGGAAAG AGGTAGGTTT 10320

GAAGTATTTG GCCAAGTATG CTGAAAAATG GCTGTTCTAA AACAGTCCCG TGGATCCGAG 10380

ATAAGGGATT AAGAAAACCA GATAAGATGA CCAGCATACT GGGAAGGATA TAGAGGAGAA 10440

AGAGACGGGG GGTGTCAGCC TGAAAATGTT TTGACTCCTG ACGAATTGTT TTTAAATCAA 10500

TTTTTGGATA GTTCATTCTC TTATTATACC ATAGTTCTTA TACATAGTTC GTGACAGTTC 10560

CTACTTTTTT TGATAAAATC ATACAGTGTG TCCTTGGGCA CACTGTATGA ACTGGGACTG 10620

TCTTTCCCAG CTTCGGAGGT AAAAAATGTC AGATTCACCA ATCAAATATC GTTTGATTAA 10680

GAAAGAAAAA CACACAGGAG CTCGTCTGGG AGAAATCATC ACTCCCCACG GTACCTTTCC 10740

GACACCTATG TTTATGCCAG TTGGGACACA AGCCACTGTC AAAACTCAGT CACCTGAAGA 10800

ATTGAAGGAG ATGGGTTCGG GAATTATCCT ATCAAACACC TATCATCTCT GGCTTCGCCC 10860

TGGAGATGAA CTCATTGCAC GCGCTGGTGG TCTCCACAAG TTCATGAATT GGGACCAGCC 10920

TATCTTGACA GATAGTGGTG GTTTTCAGGT TTATTCTTTA GCAGATAGCC GTAATATCAC 10980

AGAAGAAGGA GTAACCTTTA AAAATCATCT AAATGGTTCT AAGATGTTCC TATCCCCAGA 11040

AAAAGCCATC TCTATTCAGA ATAATCTGGG TTCAGACATC ATGATGTCCT TTGATGAATG 11100

TCCTCAGTTT TATCAACCTT ATGACTACGT TAAGAAATCG ATCGAGCGTA CCAGCCGTTG 11160

GGCTGAGCGT GGTTTGAAGG CTCACCGTCG TCCACATGAC CAAGGTTTGT TTGGAATTGT 11220

GCAAGGTGCA GGATTTGAAG ACCTTCGCCG CCAATCAGCT CATGATCTTG TCAGCATGGA 11280

TTTCTCAGGC TACTCTATCG GTGGTTTGGC AGTGGGAGAA ACCCATGAAG AGATGAATGC 11340

GGTCTTGGAC TTTACAACTC AACTGCTGCC TGAAAATAAA CCTCGTTATC TGATGGGTGT 11400

GGGAGCGCCA GATAGCTTGA TCGATGGGGT CATTCGTGGG GTGGATATGT TTGACTGTGT 11460

CTTACCGACT CGAATTGCTC GTAACGGGAC TTGTATGACC AGTCAAGGAC GTTTGGTTGT 11520

GAAAAATGCC CAGTTTGCTG AGGACTTTAC GCCACTGGAT CCTGAGTGTG ATTGCTACAC 11580

ATGTAATAAC TATACACGCG CTTACCTTCG TCACCTGCTC AAGGCTGATG AAACCTTTGG 11640

TATCCGCTTG ACTAGCTACC ACAATCTTTA CTTCTTGCTT AACCTGATGA AGCAAGTGCG 11700

ACAAGCCATC ATGGATGACA ATCTCTTGGA ATTCCGTGAG TATTTTGTGG AAAAATATGG 11760

CTATAATAAG TCAGGACGTA ATTTCTAAAA TGGAATTGAT ATAAAAAAAT CCTAAGTTTT 11820

CTCTTAGGAT TTTTCTTCTT TTTTTGATAG AATAAAGTGT ACAATGAAAG GAAGAATAAA 11880

CTCGTATGCG CATTAAATGG TTTTCCTCGA TTAGG 11915

(2) INFORMATION FOR SEQ ID NO: 97:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9069 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

GAGAGGGCAA CAGTTCTATC GCTTCAAATT TTTTCTTGGT TTGCAGATAT TCAAGAATCG 60

GGAGTTTTTC TATAGTATTC GGCAGATTTA TTACAGCCAA GCATCTCAAA AATACGGACA 120

GCATCCTCCA TCTTTTTCTG GCCTTCCTTG ACTCTACCTT GCTTGCTATC AAGGAGACCT 180

TCTGCCCACA GATAAACAAT TCGGAAATAG GTCTCATTTT CCTTGTAGAA ATGCTCTTCG 240

ATAACACGTT TAAAATAATA GGCATTGGTA AATTCTTCAC ACTCAATACT AGCTAAAAAG 300

CCATTCAATA GTATAGTATG AAAAAGGTTT CGATTGCCAG ACATTTCCAT TAGAAAATCA 360

GATTTACGTA CCATTTCTCG TACATATCTA GTAAAAAGAG AAACAGATAA AAATGGAGAA 420

CTGACTGAAA ATAAATTGAG TTCATAGATT CCCCAGATCT CGGTAGAAAA CAAATAATCA 480

TGAAGGACTT TTCCTTCCTC TGCTGTTAAG TCTACCCTTT CATCTATGCT CTTCATATAA 540

GACTTGATAA TAATGGCATT TAGAATATGT TTCTGTTTGT TGTGAGAATG GGCATGCTTT 600

TATACTCCCT GCGATATAAG TCCTCAAGAG GTGCTATATT CTTTGGTTCC AAGACATCTG 660

TAATTTCTTT TCTCAACTCA GAATCTGTAT CATACTGGAA ACCTCTTGCC AGAAAGAGGA 720

TCTCCTCCAC ACTGGCAGAT ATATTTTCCA GAGCAAATAG AAACTTTTCC ACCGAAAGCT 780

CACTCTGACC TGTTTCAAAA CGGGACAACA TAGACGGCGA AAATTGTCCT CCGGTTGCTT 840

GTCTCAGTGA GATATTTCTT GACTCTCGTA ATTGTCTAAA GACTTTTCCA ATCTGCTCCA 900

TAGACTTCCC CTTGATTCCG TATTTTCTTC ATTTTATCAT ATTTTTCAGA AAATTCATCA 960

AAAACTTGCC AAATTGTCAG AATTATGAGA AAATAGAGGA TATTTATCAC GTGGAGGGAC 1020

TGCTATGAGA GACGATATCA AAATCAATGA CCGTGCTTTG GCCTTGCAAG ACCAAATTAT 1080

CGAAAAACTA GAGAAAGTTT TTGATACAGA TGTGGAATTG GATGTTTACA ATCTAGGTCT 1140

GATTTATGAA ATCAATCTGG ATGAAACGGG GCTCTGCAAG ATTGTCATGA CCTTCACCGA 1200

TACTGCCTGT GATTGCGCCG AAAGCCTGCC TATTGAAATC GTGGCAGGTC TGAAACAAAT 1260

CGAGGGTATC AAAGATATCA AGGTTGAAGT TACCTGGTCG CCTGCTTGGA AAATCACACG 1320

AATCAGTCGC TATGGCCGTA TTGCCCTTGG ACTACCACCT CGTTAAGCAG ACCAATCACT 1380

TTTAAAGATG AAAATCAAAG GGCAAACTAG AAAACTAGCC GCAGGTTGCT CAAAACACTG 1440

TTTTGAAGTT ATGGATAGAA CTGACGAAGT CAGCTCAAAA CACTGTTTTG AGGTTGTGGA 1500

TAGAACTGAC GAAGTCAgCT CAAAACACTG TTTTGAGGTT GTGGATAGAA CTGACGAAGT 1560

CAGCCCAAAA CACTGTTTTG AGGTTGTGGA TAGAACTGAC GAAGTCAGTA ACCATACCTA 1620

CGGCAAGGCG ACGTTGACGT GATTTGAAGA GATTTTCGAG TATGAGTTTA TTTTTTACCT 1680

GACTTGTCCA TATTCCAGAA GTCTGTCACG GCTCCGCGTG AAGCAGATGA TACGATGTGG 1740

GCATATTTAC CGAGGACACC ACGGCTGTAA AGTGGTGGCA AGGTTGTTTC TGCCTTGCGT 1800

TTTTCAAGTT CTTCTTCGGA TACGGCCATA GAAATTTCTT TGGTATCTTG GTCAACCGTA 1860

ACGATATCGC CGGTACGGAG ATAGGCAATT GGTCCACCAT CCTGAGCTTC AGGAGCGATA 1920

TGTCCAACAA CCAGACCATA AGTACCACCA GAGAAACGTC CGTCCGTCAA GAGGGCCACC 1980

TTATCTCCCT GACCTTTACC AACAATCATT GAAGAAAGTG ATAGCATCTC AGGCATACCA 2040

GGACCACCTT TAGGTCCAAC AAAACGAACA ACGACTACAT CGCCATCAAC GATTTCATCT 2100

GTCAGAACGG CCTGAATCGC ATCTTCTTCT GAGTCAAAGA CCTTAGCTGG CCCAACGTGA 2160

CGACGCACTT TAACACCTGA TACCTTGGCA ACTGCACCGT CAGGAGCAAG GTTCCCGTTC 2220

AAGATGATAA GCGGACCATC CGCACGTTTT GGATTTTCAA GTGGCATGAT AACTTTTTGG 2280

CCTGGAGTCA AGTCTGCAAA GTCAGCCAAG TTTTCAGCTA CAGTCTTACC AGTACATGTG 2340

ATGCGATCTC CGTGAAGGAA ACCATTTGCC AACAAATACT TCATAACCGC AGGGACACCA 2400

CCGACTTCGT AGAGGTCTTG GAAGACATAC TGACCAGATG GTTTCAAGTC GGCCAAGTGA 2460

GGCACACGTT CTTGAATCGT ATTGAAGTCC TCAAGTGACA AGTCAACATT TGCGGCATGG 2520

GCAATGGCGA GCAAGTGAAG AGTGGCGTTT GTAGAACCAC CGAGAGCCAT CGTTACAGTG 2580

ATAGCATCTT CAAAGGCTTC ACGAGTCAAG ATATCTGATG GTTTGAGACC AAGTTCCAAC 2640

ATCTTAACAA CAGCACGTCC TGCTGCTTCG ATATCTTCTT TCTTATCAGC TGATTCAGCT 2700

GGGTGAGAGG ATGACCCTGG CAAACTCATC CCTAGAACTT CGATAGCAGT TGCCATGGTA 2760

TTAGCAGTAT ACATACCACC ACAACCACCA GGGCCAGGGC AGGCATTACA TTCAAGACGT 2820

TTCACGTCCT CAGCTGTCAT GTCACCGTGG TTCCATTTTC CGATACCTTC AAAGACAGAA 2880

ACCAAGTCGA TATCTTTACC ATCAAGATTT CCCGGTGCAA TAGTTCCACC ATAGGCGAAA 2940

ATAGCTGGGA TATCCATATT AGCAATAGCA ATCATAGATC CAGGCATGTT CTTGTCACAG 3000

CCACCGATAG CGACGAAGGC ATCCACGTTG TGACCACTCA TAGCCGCCTC GATGGAGTCC 3060

GCGATGATGT CACGAGATGT TAGAGAGAAA CGCATACCAG GCGTTCCCAT AGCGATCCCG 3120

TCCGCTACGG TAATGGTTCC AAACTGTACA GGCCAAGCGC CTGCAGATTT GACACCTTCT 3180

TTAGCCAGTT TCCCGAAATC ATGCAAGTGA ATGTTACATG GTGTATTTTC CGCCCAAGTC 3240

GAAATCACTC CCACAATCGA TGTTTCAAAG TCCTTATCTG TCATACCAGT CGCACGAAGC 3300

ATAGCACGGT TAGGTGATTT AACCATGCTG TCATAAATGC TACTGCGGTG ACGTTTATCT 3360

AATTCAGTCA TCTTATCCCT CCCATTTCAG TTTTTACTAT TATAGCACAA TTTTCGCATG 3420

AAGAACAGAA TAAAATTCTT GAATTTTCAG AAAATTCTAT ACACATGTGA AATATTTAAA 3480

ATTAAAAACA ACAAAGCGGA TTAGTGCACT TTCTGATGAC CAGAATATGC TTTTTAATCC 3540

GCTTTCTTTA AATAACGTAC TGTAATTTTT ACAGAAATTC TTTCAAATAA GTGTATTTAA 3600

CATCTATCTT GCATTATAAA TTTCTAGAAC CTTCTCTTTT ATATTCGATT CACTCAAACC 3660

ATACTCATTA AGAAGATAAT CCATTTTCCC TACTTGACCG AATCTTTCTT GAACACCCAT 3720

CCGATGAATT TTTGTTATTC CATCATCAGA GAATAATTCA CATAAAGCAC TGCCAATTCC 3780

ACCTATCTGA TTGTGGTTTT CTACAGTAAA TATAGTTTTT CCACTTAACA TTGTTTTTAT 3840

CTGTTCTGGT ATCGGTTTGA TTCTAAATAA ATCTATCACA CCTACTGAAT AACCTAATTT 3900

AGACAGTTCA TCTGCAACTC GAATACTTGG AGCAACCATT ATGCCAGAAG CAACGATTAC 3960

AAGATCTTCA CCATGCCTTA ACTCAATGTA GCCTTTAGAA AAATCTTCTC CACCTTGATA 4020

CACAGGAACT GGAGCTTTTC TAATTGTTCG AATATATTTT AGTCCTTTTA AGTCTAATGT 4080

CTGGTTCAAT ATTTCACGAA ATTGGATATC ATCAGTTGCT TCGAAAATGA TTGATTTAGG 4140

AATTAAACGT AACAATCCAA TTTCTTCAAA TGGCATATGT GTTCCACCAT TCATCTCTGC 4200

CGTTACTCCT GCATCTGATC CAATCACAGT GGCATCCAAT TGTGCGTATC CAAGAGAAAT 4260

AAATAATTGA TCAAATACTC TTCGTGAAGC AAAAGGACCA AATGTATGAA GATAAGGTCT 4320

AAACCCCTGA ATAGACAAGC CTGCTGCAAG GCCGACCATT TCTGCTTCCA TAATCCCAAC 4380

ATTCACATAA CGGTCTCCAA AGTCCTTTTC AAGATTATTA GTAGCCATCG AACTTGACAA 4440

ATCGGCTTCT AAGACTACTA TATCAGAATC ACTTTGATTA GCCTCTAAAA GGAAGTCTCT 4500

ATATACATGC CGTAATTCTT TCGTACTTCT CATCATTCTG TTTCCTCCAA TTCCTGACTT 4560

AATCTTTCTA CAACTGAAGT TAACATTTGT TTCTCCTCTA CAGTAGGGCG AAGATGATGA 4620

TTGGATTTCA TTTCTTCCAG CTCTTGAACC CCTTGACCTT TAATAGTATC TAATACAATG 4680

CACTTAGGTG ATGAATTATT TGACTGTTTT AATTGGACAA TCCCTTCATA AATTTCTCTA 4740

ATATCTGAAC CCTTGACCCT AATGGATTCA AATCCAAATG CTGAAAATTT TTCTACGAAA 4800

TCACCTGGAT TACAAATATC CTTTGTAAAA CCATCTAATT GTTTTTTGTT ATCATCAACA 4860

AATACAATTA AGTTGGATAA CTGTTGATGA GAAGCAAACT GTATAGCCTC CCAACATTGT 4920

CCCTCATTTA ACTCACCATC TCCAACAATA GCGTAAGTAT AAAAGGGACT CTTTCTTATT 4980

CTCTGACCAT ATGCAAGTCC AGTTGCAACA CTAATTCCTT GTCCTAAAGA GCCCGTTGTC 5040

ATATCTATGC CTGGCGTTAG ATTTCTATCA GGATGAGACG GTAATTTGGT TCCATTTGTA 5100

TTTAAAGAAT ATAAGAATTC TTTGTCAAAG AAACCATTCA AATAGAGTGT ACTGTATAGA 5160

GCTGGTCCTC CGTGACCTTT TGATAATATG AAATAATCTC TATCTCGTGC TGCAAATATT 5220

TCTGGAGTCA TTGGCATTAT TTCACCATAA AGCACCGCTA AAACTTCTAC GATAGACAGA 5280

CTTCCTCCGT AATGTCCGAA TCCAAGATGA TTCAATGTTC TAAGAGTATT TAATCGGATG 5340

TTAGTCGCAA ATTTTCTTAA CCCATCTTCT CTATTTTTAC TTAAAATCAT CCCTTATTCC 5400

TCCGTTGCAG ATGGCTTTTT AATAAAGGAT ACTCCAAACA TAACTGCTAG AATAAGAACA 5460

AGACCAATCA CAATGCCTGC TTGTGAGCCA AATTGATTTA ACATTCCTAA AATAATTCCT 5520

GATAGACCAA AATCTGCATC TGAGAAAGTT GATCCTTGGA AACCAAGTCC TCCCAAAACT 5580

GGCATTAAAA AGACTGGAAG AAAACTGATT AAAATACCTT GTAAAAATGC TCCAATAGTG 5640

GCTCCACGAA CACCACCAGA TGCATTCCCA ATGACACCTG CAGTCGCTCC ACAGAAGAAA 5700

TGAGGCACAA CACCTGGTAA GATAACAACC GTTCCTGAAG CAATCATAAT TACCATACTT 5760

ACTAAACCAC CAACAAAACT AGAGATAAAT CCAATTAGAA CTGCATTGGG TGCATAAGTA 5820

TAAACAATCG GACAATCCAA AGCAGGTTTT GAATTAGGTA CAAGACGCTC TGAAATACCT 5880

TTAAAGGCTG GAACAATTTC GCCCAAAATA AGGCGAACAC CTGCTAAAAT AACAAATACC 5940

CCTGCTGCAA ATTGACCTGC TAATTGTAAA GCATAAACTA GACCACTTGT ACCACTACTG 6000

ATTTCTTTTT CTATATATTC TGACCCTGCA AAGATAGCTA CAATAATGTA AATAACTGCC 6060

ATGGATAAAG TAATACTAAC AGTACTATCA CGTAAAAAAG CTAAACTCTT TGGAAATTTA 6120

ATGTCCTCTG TTGATTTTGA TTTGTCACCG ATAAGGCTAC CAGTAAAACC ACTCAACCAA 6180

TATCCCAAAG AACTGAAATG ACCTAAAGCT ACCTTGTCAT TTCCAGTTAA TTGAACCATA 6240

TATTTTTGCA CAAATGCTGG GGAAATACTC ATAATAATAC CGAGTGCTAA TCCTCCTAGT 6300

AAGATGAGAG GCAAGCTAGT AAAGCCAGCA ACTGATAAAA TGACCGCAAT CATACATGCC 6360

ATATATAGAG TGTGGTGCCC TGTTAAAAAA ATATATTTAA ATCGAGTAAA ACGAGCGATT 6420

AAGATATTGA ACACCATGCC TGCAAACATA ATCATTGCAG TAGCTGAGCC ATATGTTGTT 6480

AAAGCTACAG CTACAATTGC TTCATTATTC GGCACAACGC CAGATAAATG AAAAGCATGC 6540

TCAAACATGG TACCAAATGG ATTTAAAGAA TTTTGTACAA TTCCTGCACC ACCAGATACA 6600

ACTAAGAAAC CAACAAAGGT CTTAATTCCA CCTTTAATAA TATCAGGTAA TTTCTTCTTC 6660

TGAAGAACTA ATCCTAAGAT TGCAATTAAA GCTACTAAAA TAGCTGGTGT ACTAACAATA 6720

TCCAATATGA ACTTCATCAT GACGCTAGCC TCCTATATAA GTCCTTTTTC TTCACAAAGT 6780

TTAGTAATTA ATTCTCGTAG TTCATCCATA TCAATAATAC TATTTAAGAT ACGAACATCT 6840

CCAAGATGAC TAGCTGAATC AGCTAGATCA CGACCAACAA TCCAAATATC AGCTGCATTT 6900

GGATCTGCTC CACCTAAATC ATAATGTTCA ACTTCTACAT CCGAAACATT CAAATCACTC 6960

AATACAGATT CAATATTCAT CTGTACCATA AAACTTGAAC CTAATCCTGA ACCACAAGCT 7020

GTACCAATTT TTAACATTAT CTAATCCTCC TGTTTAATTA TCATTTTAAT GTCATCATAG 7080

TTTTTTGATG ATATTAAAGT TTGAACATGA TTTTTATCTC TTAAAATTGT TGTTAAATGT 7140

GACAAAGCCT TTAAATGACT CTCATTATCA ATGGCTGCAA TACAAATCAA CAATCTTACC 7200

TCTTGTTCTG GATTATCCAA TAAATAAATC GGTTCTTCCA AAACTAACAT TGACATTCCT 7260

ATTTCATTCA CACCTTCATC TGGCCGAGCG TGAGGAATTG CTACTCCCTT CCCTAAATTA 7320

ATAAAAGGTC CAAACTCTTC TACTTTTTGA ATCATTGCCT CAGGGTAGTT CTCAGTTATC 7380

TTATCTTGAT CCAAAAGCGG TTTAGCTGCT AAACGAATCG CCTCCTTCCA TCCTAATTTT 7440

TGCGAACTAA CCTGATAGGT TTCTTTGGTA ATAAGTTGTT CTAGCACTGG TACAATTTCC 7500

TTTCTATCAT TTTTTTGGTA AAGATAATTC TTTAACGCCA ATCTTAATTC CAATTCTTGT 7560

GTAATAATTC CATATCTTTT GACAATATTC AGGATTTGTT CAATCTCAAA ATCTCCATAC 7620

TCTAAATTCG GAAAATCTTT TAACACTAGT TCTACTAGTT GTATTGCTTG CTCTTCAGTC 7680

ATCATAACCG AAACTAGATA ATTTGGCTTT TCTGTCTCCA CCTTTATGGT AGAAAAAACC 7740

ATATCATAGT CACTACTAGC TTTCACCTGT AAATCATCAA TCTTTGAGGT TCCTATAAAC 7800

TCAATTTGAG GAAATAATGC TAATAGATTC TCTTTTAACA TCAATGAAGA ACTAACACCA 7860

TTAGGACAAA TGATTGCTGC TTTATACCAT TTTTGAGGCA AAGTATCTGC TTTCTTTAAA 7920

TAACCTCCGA AATGGATAAC AAAATATGCT GTTTCACTAT CAGGTATGGG ATTGTCAATA 7980

GCGTCCATCA AGGGCATCAA AGAATCTTTG ACTAATTCAA ATAAATCAGG ATAATGTTCT 8040

TTAACATGCA ATACATATTC ATTTGAACTA GGTAGGCCGA ACTTTAATCT ATAGTAAGCC 8100

GGTATAAGGT GGCGGCGAAG ATTTTCTCTC AATCCTTCCC TTTGTTTAAA ATGTAACAAA 8160

GAAATATCTT CCATTCTACT TATAATAGCC TCTGTTAATT GATTAAAGTA AACCGGAGCA 8220

ACATCTACTT CACCTTCAAA GCAACTTGAT AATAAAACGG TGATATAGCG ATAATCATCC 8280

TCAGAAAACA CCGTATCTAT AATTCCCAAA TCAACCACTG TATCCAATAA AATAGTGGTT 8340

ATATCTTGAA TAACAGGAGA TACTAATGTC TCTGAAAGAC ATACTCTTTC AACATCCCTT 8400

TGATACCTAC ACAGAATGAA TACTAAACCG AAAAGGTAAA CTTTTAATTG ATTAACAATA 8460

GGTACTAGCT GTAGCTTCTC ATAATAATCT TTAACTACCT GATCAATCAA ATCATAAGTT 8520

AATGAATACC CCCAACTGGA TAAAACATAA TCCAAACCCC AAATCCCTAT GGAGGATTCC 8580

AGCAACTCAC TAACCATTTG AAAAGCTAAG CGGTGCTTAT TCCACTCTGA ACCGTGTAAA 8640

GTATAACCTT TTGCTCTACT GTACCCTAGC TCCAAATCAT TATCTAACAT AATCTTTCTT 8700

AATGATTGAA TATCAGATAA GGTTGTATTC TTACTTACTT TCAAAAAGTC TTGGTAATGA 8760

CTATTCGATA TAAAATCTAA TCGGCAAAAA GTGTAAAGAT AGATTAAAGC TAAGCGAGTC 8820

GACTTTGGTA AAACCAATTC ATCCGACTTA ATAATATCTG TCAAAGACTG CTTCGTACGA 8880

TTTGATAAAC TATAGCGACC TTGCTTTTTA TCCAGCACTA TCCCTTTATT AGCTAGATAA 8940

GGCACTAAAT AATCTATTCC TTCTTTGACT TCCTTTATAG GTAAGCTCAC CTTAACAGAT 9000

AATTCATATA ACGATAGCTC ACAATGATCC ATCAAAGTCA TCAAAATAAC TAGTGCTCTA 9060

TAATCAAAC 9069

(2) INFORMATION FOR SEQ ID NO: 98:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8654 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

CGAGACAACA AGATGAAGAA AAATTTGCCC TATCGTTTGT GGCGCTTGCA AGTGTAGCAC 60

TTCTTGCAGC CTGTGGAGAA GTGAAGTCTG GAGCAGTCAA CACTGCTGGT AACTCAGTAG 120

AGGAAAAGAC AATTAAAATC GGGTTTAACT TTGAAGAATC AGGTTCTTTA GCTGCATACG 180

GAACAGCTGA ACAAAAAGGT GCCCAATTGG CTGTTGATGA AATCAATGCC GCAGTGGTAT 240

CGATGGAAAA CAAATCGAAG TAGTCGATAA AGATAATAAG TCTGAAACAG CTGAGGCTGC 300

TTCAGTTACA ACTAACCTTG TAACCCAATC TAAAGTATCA GCAGTCGTAG GACCTGCGAC 360

ATCTGGTGCG ACTGCAGCTG CGGTAGCGAA CGCTACAAAA GCAGGTGTTC CATTGATCTC 420

ACCAAGTGCG ACTCAAGATG GATTGACTAA AGGTCAAGAT TACCTCTTTA TTGGAACTTT 480

CCAAGATAGC TTCCAAGGAA AAATTATCTC AAACTATGTT TCTGAAAAAT TAAATGCTAA 540

GAAAGTTGTT CTTTACACTG ACAATGCCAG TGACTATGCT AAAGGGATTG CAAAATCTTT 600

CCGCGAGTCA TACAAGGGTG AAATCGTTGC AGATGAAACT TTCGTAGCAG GTGACACAGA 660

CTTCCAAGCA GCCCTTACAA AAATGAAAGG GAAAGACTTT GATGCTATCG TTGTTCCTGG 720

TTACTATAAT GAGGCTGGTA AAATTGTAAA CCAAGCGCGT GGCATGGGAA TTGACAAACC 780

AATCGTTGGT GGTGATGGAT TCAACGGTGA GGAGTTTGTA CAACAAGCAA CTGCTGAAAA 840

AGCATCAAAC ATCTACTTTA TCTCAGGCTT CTCAACTACT GTAGAAGTTT CAGCTAAAGC 900

TAAAGCCTTC CTTGACGCTT ACCGTGCTAA GTACAATGAA GAGCCTTCAA CATTTGCAGC 960

CTTGGCTTAT GATTCAGTTC ACCTTGTAGC AAACGCAGCA AAAGGTGCTA AAAATTCAGG 1020

TGAAATCAAG AATAACCTTG CTAAAACAAA AGATTTTGAA GGTGTAACTG GTCAAACAAG 1080

CTTCGATGCA GACCACAACA CAGTCAAAAC TGCTTACATG ATGACCATGA ACAATGGTAA 1140

AGTTGAAGCA GCAGAAGTTG TAAAACCATA ATAGAAAAAT GTTGAAATAG GGAATGAGCC 1200

TTTGACTCAC TCCCTGTTTC GATATTTAAT ACTCTTCGAA AATCTCTTCA AACTGCGTCA 1260

ACGTCGCCTT GGATTATATA TGTGACTGAC TTCGTCAGTC TTATCTACAA CCTCAAAGCA 1320

GTGCTTTGAG CAACCTGCGG CTAGTTTCCT AGTTTGCTCT TTGATTTTCA TTGAGTATAA 1380

GAACCTATCA AAAAGTGAGG GAAAACCCTC GGAATTATAA ATAGAAAGAG TGAATCTTAT 1440

GCTCCAACAA CTCGTAAATG GTTTGATTCT AGGTAGTGTT TACGCGCTGT TAGCCCTAGG 1500

ATATACCATG GTTTACGGAA TTATCAAGCT CATCAACTTC GCCCATGGTG ATATTTATAT 1560

GATGGGAGCC TTTATCGGTT ATTTCTTGAT CAATTCTTTC CAAATGAATT TCTTTGTAGC 1620

GCTTATTGTA GCTATGCTAG CGACAGCTAT TCTTGGTGTC GTGATTGAGT TTCTTGCTTA 1680

CCGACCTTTG CGCCACTCTA CTCGTATTGC TGTTTTGATT ACGGCTATTG GGGTTTCTTT 1740

CCTATTGGAG TATGGAATGG TCTATCTGGT TGGTGCCAAT ACCCGTGCCT TCCCTCAAGC 1800

GATTCAAACA GTTCGATATG ATTTGGGACC AATTAGCTTA ACAAATGTGC AGTTAATGAT 1860

TTTGGCCATT TCCTTGATTT TGATGATTTT GTTACAAGTC ATTGTCCAAA AGACTAAGAT 1920

GGGGAAAGCC ATGCGTGCAG TATCAGTAGA TAGCGACGCG GCGCAATTGA TGGGGATCAA 1980

TGTAAACCGT ACGATTAGCT TTACCTTCGC TTTGGGTTCT GCTCTTGCGG GTGCGGCTGG 2040

TGTTCTGATT GCTCTTTATT ATAACTCTCT TGAGCCTTTG ATGGGGGTTA CTCCAGGTCT 2100

TAAATCTTTC GTTGCCGCAG TACTTGGTGG TATCGGAATT ATTCCTGGTG CGGCTCTTGG 2160

TGGCTTTGTG ATTGGTCTAT TGGAAACCTT TGCGACTGCC TTTGGGATGT CAGATTTCCG 2220

TGATGCCATT GTTTATGGAA TCTTGTTGTT GATCTTGATT GTCCGCCCAG CTGGTATCCT 2280

TGGTAAGAAT GTGAAAGAGA AGGTGTAAAC GATGAAGGAA AATTTAAAAG TTAATATTCT 2340

ATGGTTACTC CTTTTGTTAG CTGGCTATAG CTTGATTAGT GTACTGGTTT CAGTCGGAGT 2400

ACTTAATCTA TTCTATGTAC AGATTTTACA ACAAATTGGA ATTAATATTA TTTTGGCTGT 2460

TGGTCTCAAC TTAATCGTTG GTTTTTCAGG ACAATTTTCA CTTGGTCATG CTGGTTTCAT 2520

GGCGATTGGT GCCTATGCAG CAGCTATTAT TGGTTCTAAA TCACCAACCT ACGGTGCCTT 2580

CTTTGGAGCT ATGCTTGTAG GGGCTTTGCT TTCAGGAGCA GTTGCCTTAC TTGTCGGCAT 2640

TCCAACCTTG CGCTTGAAGG GGGACTATCT TGCGGTAGCA ACTCTGGGTG TTTCTGAAAT 2700

TATCCGTATC TTTATCATCA ATGGTGGAAG CCTTACAAAT GGTGCGGCAG GTATCTTAGG 2760

GATTCCTAAC TTTACAACTT GGCAAATGGT TTACTTCTTT GTCGTGATTA CAACCATTGC 2820

AACCTTGAAC TTCTTGCGTA GCCCAATTGG TCGTTCAACC CTCTCTGTTC GTGAAGATGA 2880

AATCGCTGCT GAGTCAGTTG GGGTTAATAC GACTAAAATT AAAATCATCG CTTTTGTCTT 2940

TGGTGCCATT ACTGCAAGTA TTGCTGGGTC ACTTCAGGCA GGATTTATCG GGTCTGTTGT 3000

ACCGAAAGAT TACACCTTCA TCAACTCAAT CAACGTTTTG ATTATTGTTG TATTTGGTGG 3060

ACTCGGTTCC ATTACAGGTG CGATTGTTTC GGCTATTGTT CTGGGAATTT TGAATATGCT 3120

TCTCCAAGAT GTTGCTAGTG TGCGTATGAT TATTTACGCT TTGGCCTTGG TATTGGTAAT 3180

GATTTTCAGA CCAGGTGGAC TCCTTGGAAC ATGGGAACTG AGCCTATCAC GTTTCTTTAA 3240

AAAATCTAAG AAGGAGGAAC AAAACTAATG GCATTACTTG AAGTAAAACA GTTAACCAAA 3300

CATTTTGGTG GTCTAACAGC TGTTGGAGAT GTGACTCTTG AATTGAACGA AGGGGAACTG 3360

GTTGGATTAA TCGGTCCAAA CGGAGCTGGG AAAACCACCC TTTTCAACCT TTTGACCGGT 3420

GTTTATGAAC CAAGCGAGGG AACAGTAACC CTAGATGGTC ACCTTTTGAA TGGGAAATCA 3480

CCTTATAAGA TTGCCTCTTT GGGACTTGGA CGTACTTTCC AAAATATCCG TCTCTTTAAA 3540

GATTTAACAG TTTTAGATAA TGTTTTGATT GCTTTTGGAA ACCATCACAA ACAGCATGTT 3600

TTTACTAGTT TCTTACGCTT ACCAGCTTTT TACAAGAGTG AAAAAGAATT AAAGGCTAAA 3660

GCTTTGGAAT TGTTGAAAAT CTTTGATTTA GATGGTGATG CAGAGACTCT TGCTAAAAAT 3720

CTTTCCTACG GACAACAACG TCGTTTGGAA ATTGTTCGTG CCCTTGCTAC GGAACCTAAA 3780

ATTCTCTTCT TAGATGAACC AGCAGCAGGT ATGAACCCAC AGGAAACAGC CGAATTGACT 3840

GAGTTAATTC GTCGTATCAA AGATGAGTTT AAGATTACAA TCATGTTGAT TGAACACGAT 3900

ATGAATCTGG TCATGGAAGT AACAGAACGT ATCTACGTAC TTGAATATGG CCGTTTAATC 3960

GCTCAAGGAA CTCCAGACGA AATTAAGACC AATAAACGCG TTATCGAAGC TTATCTAGGA 4020

GGTGAAGCCT AATGTCTATG TTAAAAGTTG AAAATCTTTC TGTGCATTAC GGTATGATCC 4080

AAGCAGTTCG TGATGTAAGC TTTGAAGTTA ATGAAGGAGA AGTTGTTTCC CTTATCGGTG 4140

CCAACGGTGC AGGTAAGACA ACTATTCTTC GCACCTTGTC AGGTTTGGTT CGACCAAGTT 4200

CAGGAAAGAT TGAATTTTTA GGTCAAGAAA TCCAAAAAAT GCCAGCTCAG AAAATCGTGG 4260

CAAGTGGTCT TTCACAAGTT CCAGAAGGAC GCCACGTCTT TCCTGGCTTG ACTGTTATGG 4320

AAAATCTTGA AATGGGAGCT TTCTTAAAGA AAAATCGTGA AGAAAATCAA GCTAACTTGA 4380

AGAAGGTTTT CTCACGCTTT CCTCGTCTTG AAGAACGGAA GAACCAAGAT GCAGCCACTC 4440

TTTCAGGGGG GGAACAACAA ATGCTTGCCA TGGGACGCGC CCTCATGTCA ACACCAAAAC 4500

TTCTTCTTTT AGATGAACCA TCAATGGGAC TTGCCCCAAT CTTTATCCAA GAAATTTTTG 4560

ATATCATTCA AGATATTCAG AAGCAAGGAA CAACGGTCCT CTTGATTGAA CAAAATGCCA 4620

ATAAAGCACT TGCAATCTCT GACCGAGGAT ATGTACTGGA AACAGGGAGA ATCGTCCTAT 4680

CAGGAACAGG AAAAGAACTC GCTTCATCAG AAGAAGTCAG AAAAGCATAT CTAGGTGGCT 4740

AAAACAATCC AGTGGATTGT TTTAGTCGGC AGATGGAGAT TACGAAGTAA TCATCAATAT 4800

AGTCCGGGGG ACCTTTTTAG TCGGTAGATT GAGATTGCAA ACAAATCTGC ATCTACATTG 4860

AAAGCTTAAT TTCTAATAAT TGAAAAAATC GAATGAAAAA TTTCTTACCT TCATTCACAG 4920

AGCTCGATTT CAGAGCTCTT TTTGCTAGCT TATTCATACT TTTCTGAATT TCGAAAAAGA 4980

AATGTAAGCG TTTGATAGAT TTACAAAAAG ATTGTATAAT AGGGATAAGA ATAGAAAAGG 5040

AGAAGTCTCA TGGCAGTTAA AGATTTTATG ACCCGCAAGG TAGTTTATAT TAGTCCAGAT 5100

ATAACAGTAT CTCATGCAGC AGATTTGATG AGAGAGCAAG GTTTGCACCG TCTGCCTGTT 5160

ATCGAAAATG ATCAATTAGT TGGTTTGGTG ACTGAGGGAA CCATTGCACA AGCAAGTCCA 5220

TCTAAAGCAA CAAGTCTTTC TATCTATGAG ATGAATTATC TTCTGAATAA GACAAAAGTA 5280

AAAGATGTCA TGATTCGCGA TGTTGTCACT GTCTCAGGCT ATGCTAGTCT AGAAGATGCA 5340

ACTTATCTGA TGTTGAAAAA TAAGATTAGT ATTCTCCCTG TCGTAGATAA CCATCAAGTA 5400

TACGGAGTTA TTACTGACCG TGACGTTTTC CAAGCCTTTC TTGAAATTGC AGGTTATGGC 5460

GAAGAAGGGA TTCGTGTACG CTTTGTTACA GAAGATGAAG TTGGTGTTCT TGGAAAAATT 5520

GTTTCTTTGA TTGTAGAAGA AAATTTGAAT ATCTCCCATA CAGTCAATAT TCCGCGTAAG 5580

GATGGTAAGG TGATTATCGA AGTGCAAATC GATGGATCAA TTGATTTACC AGCCTTGAAA 5640

GAAAAATTTG AAGCAAATGG TATTCAAGTG GAAGAAATCG CTCGCACTTC AGCAAAAGTC 5700

TTGTAAGAAG GGAAGCCCAA AGGCTTCTTT TTTCATGAAA AGGGGATTAG AGCAAAAGAT 5760

GGAAAGAAAT GATAAAATAT GCTATAATGA AATAATGTAA AAAAGGAGTA TTTATGGACA 5820

TTTCAGTAAT TCGTCAGAAA ATTGACGCAA ATCGTGAAAA ATTAGCTTCT TTCAGGGGGT 5880

CTCTTTGACC TCGAAGGGCT AGAGGAAGAG ATTGCCATCT TGGAAAACAA GATGACAGAA 5940

CCTGATTTTT GGAACGATAA TATTGCGGCC CAAAAAACGT CGCAAGAATT AAATGAATTA 6000

AAAAACACTT ACAATACCTT CCATAAGATG GAAGAGTTGC AGGATGAAGT CGAAATTTTA 6060

TTGGATTTTT TGGCTGAAGA CGAGTCAGTG CATGATGAAC TGGTAGCGCA GTTAGCCGAA 6120

CTTGATAAGA TAATGACCAG CTACGAGATG ACTCTACTCT TGTCAGAACC TTATGACCAC 6180

AACAATGCCA TCTTGGAAAT CCATCCAGGT TCTGGTGGTA CTGAGGCGCA GGACTGGGGT 6240

GATATGTTGC TTCGTATGTA TACTCGTTAT GGTAATGCTA AAGGCTTTAA AGTGGAAGTG 6300

TTGGATTACC AAGCAGGTGA TGAGGCTGGT ATTAAGTCGG TAACTTTATC ATTTGAAGGG 6360

CCTAATGCCT ATGGTCTCCT CAAGTCAGAA ATGGGTGTTC ACCGCTTAGT GCGAATCTCA 6420

CCATTTGACT CTGCCAAACG TCGCCATACC TCTTTCACAT CTGTAGAAGT GATGCCAGAA 6480

TTGGATGATA CTATTGAAGT GGAAATCCGT GAAGATGATA TCAAGATGGA TACCTTCCGT 6540

TCAGGTGGTG CCGGTGGACA AAACGTCAAT AAGGTTTCAA CAGGTGTACG TTTAACCCAC 6600

ATTCCAACTG GAATTGTTGT CCAATCAACA GTAGATCGTA CCCAGTATGG AAATAGAGAT 6660

CGTGCCATGA AGATGTTGCA GGCTAAGCTC TATCAAATGG AGCAAGATAA GAAGGCTGCG 6720

GAGGTAGATT CTCTCAAAGG TGAGAAAAAG GAGATCACTT GGGGAAGCCA AATCCGTTCT 6780

TATGTCTTCA CGCCTTATAC TATGGTAAAA GATCACCGAA CTAGCTTTGA GGTTGCTCAG 6840

GTAGATAAGG TTATGGATGG GGACCTAGAT GGTTTTATCG ATGCTTATCT CAAGTGGCGA 6900

ATTAGCTAAG ATAGAAAGGA ACTCACATGT CAATTATTGA AATGAGAGAT GTCGTTAAAA 6960

AATACGACAA CGGAACAACT GCTCTACGCG GTGTTTCGGT TAGCGTTCAA CCGGGGGAAT 7020

TTGCTTACAT CGTAGGACCT TCAGGAGCAG GGAAGTCAAC TTTTATTCGT TCTCTGTATC 7080

GTGAAGTAAA AATCGATAAA GGAAGCCTAT CAGTTGCTGG TTTTAATCTG GTTAAGATCA 7140

AAAAGAAAGA TGTCCCGCTT CTACGTCGTA GTGTTGGGGT TGTCTTCCAG GATTATAAAT 7200

TGTTACCAAA GAAAACTGTC TATGAAAATA TTGCTTACGC TATGGAAGTA ATCGGGGAAA 7260

ATCGCCGTAA TATCAAAAGA CGAGTGATGG AAGTTTTGGA CTTGGTTGGA TTGAAGCATA 7320

AGGTTCGTTC TTTCCCAAAT GAACTCTCAG GTGGGGAGCA ACAGCGGATT GCGATTGCGC 7380

GTGCAATTGT AAATAATCCC AAAGTATTGA TAGCTGATGA GCCAACAGGA AATCTGGATC 7440

CGGATAATTC ATGGGAAATT ATGAATCTCT TGGAACGGAT TAACyTACAA GGAACAACTA 7500

TTTTGATGGC GACTCATAAT AGCCAGATTG TAAATACCTT GCGCCACCGT GTCATTGCCA 7560

TTGAAAATGG CCGTGTCGTT CGTGACGAAT CAAAAGGAGA GTATGGATAC GATGATTAGT 7620

AGATTTTTTC GCCATTTATT TGAAGCCTTA AAAAGTTTGA AACGAAATGG TTGGATGACA 7680

GTAGCTGCTG TCAGTTCAGT CATGATTACT TTGACCTTGG TGGCAATATT TGCATCTGTT 7740

ATTTTCAATA CAGCGAAACT AGCTACAGAT ATTGAAAATA ATGTCCGTGT AGTAGTTTAT 7800

ATCCGAAAGG ATGTGGAAGA TAATAGTCAG ACAATTGAAA AAGAAGGTCA AACTGTTACA 7860

AATAATGACT ACCACAAGGT ATATGATTCT TTGAAGAACA TGTCTACGGT TAAAAGTGTT 7920

ACCTTTTCAA GTAAAGAAGA ACAATATGAA AAATTAACCG AGATAATGGG AGATAACTGG 7980

AAAATCTTTG AAGGAGATGC CAATCCTCTC TATGATGCCT ATATTGTAGA GGCAAACACT 8040

CCAAATGATG TAAAAACTAT AGCCGAAGAT GCTAAAAAAA TTGAAGGTGT CTCTGAGGTT 8100

CAAGATGGCG GTGCCAATAC AGAAAGACTC TTCAAGTTAG CTTCATTTAT CCGTGTTTGG 8160

GGACTAGGGA TTGCTGCTTT GTTAATTTTT ATCGCAGTTT TCTTGATTTC AAATACCATT 8220

CGTATTACCA TTATTTCCCG CAGTCGCGAA ATTCAAATCA TGCGCTTGGT CGGAGCTAAA 8280

AACAGTTATA TCCGTGGACC GTTCTTGTTA GAAGGAGCCT TTATCGGTTT ATTGGGAGCT 8340

ATCGCACCAT CTGTTTTGGT CTTTATTGTT TATCAAATTG TTTACCAATC TGTCAACAAA 8400

TCGTTGGTAG GGCAAAATCT ATCCATGATT AGTCCAGATT TATTTAGTCC GTTGATGATT 8460

GCCCTACTAT TTGTGATTGG GGTTTTCATT GGTTCATTGG GATCAGGAAT ATCCATGCGC 8520

CGATTCTTGA AGATTTAGGT AAAATAGCTG CTTTTATGAG GAGATTGTAA AATCTCCTTT 8580

TTTGCTACAA GAGTTTTTGA AAAGAGATGC GCAGAAGAAA AGAGCTTCCA AAGAAGTCCC 8640

CCAGAGAAGA CTTC 8654

(2) INFORMATION FOR SEQ ID NO: 99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19718 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

TGTCGCGTCA AAATCATTAC TATGGCTATG TATAGCCCTT ACTATGACTT GGCTAAACAC 60

GTTCGCTTTC AAATTTCTAG GCTCAGGCTG AAACAGTCTC CCAGGCTGTT CACTCCCGAA 120

TGCTAAAATC GTTCTTGATC GCTTTCACAT TGTACAACAT CTTAGCCGTG CTATGAGTCG 180

TGTGCATGTC CAAATCATGA ATCAGTTTCA TCGAAAATCC CATGAATACA AGGCTATCAA 240

GCGCTACTGG AAACTCATTC AACAGGATAG CCGTAAACTG AGTGATAAGC GATTTTATCG 300

CCCTACTTTT CGCATGCACT TAACAAATAA AGAAATTCTT GACAAGATTT TAAGCTATTC 360

AGAAGACTTG AAACACCACT ATCAGATCTA TCAACTCTTA CTTTTTCACT TTCAGAACAA 420

AGACCCTGAG AAATTTTTCG GACTCATTGA GGACAATCTG AAGCAGGTTC ATCCTCTTTT 480

TCAGACTGTC TTTAAAACCT TTCTCAAAGA TAAAGAAAAG ATTATCAACG CCCTTCAACT 540

ACACTATTCT AATGCCAAAC TGGAAGCGAC CAATAATCTC ATCAAACTTA TCAAGCGCAA 600

TGCCTTTGGT TTTCGAAACT TTGAAAACTT CAAAAAACGG ATTTTTATCG CTTTGAACAT 660

CAAAAAAGAA AGGACGAAAT TTGTCCTTTC TCGAGCTTAG CTGACTTCAA CCCACTACAG 720

TTGACAAAGA GCCTAATTTC CATAAAAATT GACATGGAAA TTATAAAACC ATTACTAGTT 780

TAGTCCTTTT TGATAACGTG CCAATTCGGC TTGGTTCGCC CAAACATAGT GACCTGGACG 840

GATTTCTACC ATAGATGGCT TATCAGTCTC ATAGTCGTGT TGACTTGGAT CGTAAACCTT 900

CAAGACCTTC TTACGTTCCA AGATTGGATC TGGGATTGGT ACCGCTGAAA GCAAGGCTTG 960

AGTATATGGG TGAATTGGAT TGTTAAACAA TTCTTCTGTT TCTGCAACCT CTACAATAAC 1020

ACCCTTGTAA ATAACTGCGA TACGATCTGA AATAAAGCGA ACAACCGACA AGTCATGGGC 1080

GATGAAGAGA TAGGTCAGGC CGAGCTCTTT TTGGAATTTT TTGAGCAAGT TCAAGACTTG 1140

GGCACGTACA GAAACGTCCA AGGCTGAAAT TGGCTCATCT GCAATAACAA AGTCTGGTTG 1200

CATGACCAAG GCACGGGCAA TACCGATACG TTGACGTTGA CCGCCTGAGA ATTCATGAGG 1260

GTAACGAGTC AAGTGCTCAG CAAGAAGACC TACTTCACGG ATAATATTTT GAACTTTCTC 1320

TTTACGTTCT TCTTCATCCT TAAATAAACG GTGATTGTAA AGACCTTCAG AAATAATATA 1380

ATCAACAGTC GCACGTTCAT TCAAACTTGC GGCAGGGTCT TGGAAAATCA TCTGGATTCG 1440

ACGAATCAAT TCCGCAGCTT GTTCACGCGA TTTCTTACCA TTAATCTTTT GACCATCAAA 1500

AATGATATCT CCATTACTTG TATCATTTAG ACCGATGATA GCACGACCAA TAGTTGTTTT 1560

CCCACTACCG GACTCACCTA CAAGCGAGAA AGTTTCTCCC TTGTTGATAA AGAAGTTAGC 1620

ATTTTTAACC GCGACAAACT TCTTACTTCC TTCACCGAAG GAAATTTCTA AATCTTTGAT 1680

TTCTACTAAT TTTTCAGACA TTTCCTTCCT CCTAGTCAGC CAGATGGGCA AATCCCATTT 1740

TTTCACGGAT CTTATCATGG AGATTTGCAA TCACAGCTGG TTTTTCTACT TTCGGAGCAT 1800

CCTCATGAAG AAGCCAAGTT TTAGCCCAAT GTGTCTCTGA TACTGAGAAT TGAGGAGCTT 1860

TTTGTTCGAA GTCAATCTGC ATTGCGTAGT CAGAACGCAA GGCAAAAGCA TCCCCTTTCA 1920

GGTCAGTATA AAGTGACGGA GGTGTTCCTG GGATTGAGTA AAGATCCCCT TTATCATCAG 1980

CAAGCTGAGG CAAGCTAGAC AAGAGACTCC ATGTATATGG ATGGCGAGGG TCATAGAAGA 2040

CTTCCTCAAC CGTTCCATAC TCAACGATTT CTCCTGCATA CATAACCGCT ACCTTATCCG 2100

CAATACTTGC CACCACACCA AGGTCGTGGG TAATAAAGAT TGTTGTGAAA TGATACTCGT 2160

TTTGTAAAGA TTTTAGCAAA TCAATAATCT GAGCTTGAAT AGTTACATCC AAGGCAGTTG 2220

TTGGCTCATC ACAGATCAAG ACATCAGGTC GGCAGGCAAG GGCAATAGCA ATAACGATAC 2280

GTTGACGCAT TCCTCCAGAA TATTGGAATG GGTATTCATT AAAACGTCTA TCTGCGTCTG 2340

GAATGCCAAC CTTATTCATG TAGTCAATGG CCAATTCTTT CGCTTCTTTA GCTGTTTTTC 2400

CTTGGTGTTT TACAATAACT TCTGTAATCT GACTACCAAT TGTTTTAATG GGGTCCAAAC 2460

TAGTCATTGG GTCCTGGAAG ATAGTCGCAA TCTTAGCACC ACGAATTTGT TCCCAATCCT 2520

TGTGAGAAGA TAAAGCTGTC AAGTCCTGAC CACGGTAGTC AATACTACCT TGGGCAATAC 2580

GACCATTTTC TTCGAGCATA CCTGTGAAGG TCTTTGTCAA AACAGATTTA CCTGATCCTG 2640

ACTCACCTAC CAAGGCTAAT ACTTCTCCTT CGACTAGTTC AAGGGAAACG CCGCGAATGG 2700

CTGTCAATAC TTTGTCACGA ACGTCAAATT CCACGACAAT ATCGCGAGCA GTCAAAATTA 2760

CATTTTTTTC TTTTGTCATT TCTACTCCTA TCTATGTGTA CGTGGATCAC TAGCATCCGC 2820

TAAGTTTTGA CCAACTACGA AAAGGGACAA GGATACCAAG ACAAGGGTTG TCAATGGAAT 2880

CCAGAACAAG TAAGCATTGG TTGTTACGTT TTGTGAATAA TCCGAAATCA AACGACCCAA 2940

ACTTGGCACT GTAATCGGTA ATCCAAGACC GAAGAAAGAC AAGAAGGCTT CGTATGAGAT 3000

AAAGCTTGGA AGCATTTGAG TCATGGTTGT CACAATAACA GATACCAATT GAGGCATGAT 3060

ATTTTTGGCA ACAATCTTCA AGGTTGGTGT TCCCAAAGTA CGTGACGCCA AGTTGTATTC 3120

CAAGTCACGA TAGCGCAAGA TTTGCACACG GATCATGAAG GCAATACCAA TCCATGTTGT 3180

TACGCTCATG GCAAAAATCA GATTCCAGAA TCCAGCTCCG ATTGAGTAAG TCAAGACAAT 3240

AACAATCAAA AGAGGTGGGA TGTTTGAGAT GACGTTGTAA ACTTCCATCA TGACACGGTC 3300

AACTGATTTT GAAATACCCC AAATACCACC GACAAAAACA CCGATAACCA AGTTAATCAC 3360

TGTCGCAATC ACAGAAATGA GGATGGAGTT ACGAGCTCCG AACCAGACAC CGTCAAAGAG 3420

CGATTTACCG TTACTGTCAG TACCGAACCA ATGCTCCGCA TTTGGCTTGA TATAACGAAC 3480

ACTAAAGTCG TTTACCTTGC TGACATCATT GAAATCAAAC TTAGAAAACA TTGGGTAGAT 3540

GAAACTTATC AAAATGATGG CTACCAAGAT TCCCAACATG ACTACAGTTG ATTTTTTCTT 3600

CATAAATTGT TTAAACACTG ATTTCCAGTA AGAATATGCT GGCGCATCAA TAGTTTCAGA 3660

GGCAAAATCG TCACGTTTTA CAAACTGAAA TTTTTCTTTA TCGATTGTAG ACATTATTTG 3720

CCTCCTTTCT CAGTCAATTT AATACGTGGG TCAATAATAG TCATCCAAAT ATCTCCCAAA 3780

AGACGTGAGA AGATAGAAAT ACATGTAAAG ATGAAGACAA GACCAACGAC CATAGAGTTA 3840

TTAGATGCTT TTACAGAGTC AATCAACATT TTACCCATAC CTGGGAAGGC GAAGACTGTT 3900

TCAGTAAGGG TTGCACCACC GATAACCCCA ATAATGGCAG CAGGAATTCC TGAAACCAGC 3960

GGAACCATGG CATTTTTAAA GATGTGTTTG TTTGAAATTT CTTTTTCAGA CAAACCTTTT 4020

GCACGAGCGA AACGAACAAA GTCTTGAGAT TGCAAGTCAA TCATGTAACG ACGAATCCAA 4080

ATGGCTGTAC CAGGAGCACC CAACAAACCA AGGATGACTG CTGGTAAAAC GTAAGAACGC 4140

CAATCTCCAG CTCCCAAGAT AGGGAATGAA TCTGGAAGGG CAATAGATGA TCCAATCAAT 4200

CGAACGATGT AAACCAAGGC AATCGTTGGA AGAGCAAGCA AGAAGGTCAA AGCCCCTGTT 4260

GAGAGGCTAT CAATCCAAGT GTTCTTGAAA CGAGCCATGG CTGAACCAAG TGGCACGGCA 4320

AGAGCATAGG CAAGAACCAA ACCAATCAAA CCAGTAATAG CAGAGCTGAC AATCATAGAT 4380

GGATATTGGT AATTACTTTC AGTCGCTGTA TAAGGATCAT CTTTCCCATA GCTAGCTACT 4440

TCACGAGAGT CAGCCTGACT AGGTGACTTG TAGGTTCTTG AGTAAATATT TACAGAAGAC 4500

GTTTTCTTAC CTGTTGGGAA CTGAACTTGG GCAGTTTTGG TTTGTCCTTG ACCTTGAGTA 4560

ATAACCTGAA GAACTGGTGT ATTAGCATAG GTTGGGTAAG AGTCACCTAA ATTCAAGTTC 4620

ACAAAGTTTT GATGAACAAA TGGGAACTGA CTGTTAAAGT ACAAGAGATA TTTATGTTTA 4680

GTTCCTGAAC CGACCAATGA CCATCCGATA GCTGGATCAT TTTCAAAACG AAGGTAGCGT 4740

TTCAAGTCTG GATTTTCAGG GTCTTGGATT TTATTTGTAT GGTCAATGTC AATCAAGTTA 4800

GCATAGAAGT GAAAAACACG TTCAAAAATT GGAATTTCAC GAGTAGCATA GAATTGACCA 4860

CTTTCAGTAA ATTCTCCCAA AGTCCAACCA TGACCTAATT GATTGATGTA CTTTTCATAA 4920

ATAGCTTTAT TGGTCGCATT TGCTTCTACT GTTACAGAAG AATCCATGCT ACTTGCCTTT 4980

TCTTGCAACT CTTTAGTATC GTAATACTCA ATGTAGCCCA TACGCTCAAA CACAGTATTT 5040

TCATAGTTAT CACGTTTATC AGCCGTTGTC GCAATTTTAT TATAGTTAGG ATCCTGCTTG 5100

AAAATCAATT TTCGAGGAAC CAAGGTATAG ATAATCGTGT AGGTCAAAGT CGTTACTAAG 5160

AAAATCGAAA CCAATGACCG CAAAACACGC ATAAAAATAT ATTTTTTCAT ATTATTTCCT 5220

TTAAAAATCC CAAAAGAACC TTCTCCTCAT GGAGAGAAAG TTCTATTAGA AATTATTTAC 5280

TTCACATGAC TTGCCAATTC TTTTTGAGCT TTCTCATTTG ATTCAGCTTT TTCTTTCAAC 5340

CATTTTTCAC GAGCTTTTTC ATACTCTTCC TTAGTCACCA CTTTATCTTG TGATTTCAAA 5400

TATTTGAAGT AAACATCTGA CCCCTTAGAG CCTGTTTGCG CAGAAGCTCC AGTAAATGGA 5460

ACAATTCGTG AAAGCACTGG TGCTGCACCA GAAGAAGCCA TAGCAGGAAT AAAGAGTGAA 5520

CTATCTGTCA ACCATGCTTG AGCCGCTGCA TATTTTTCAT AACGGACATT CAAGTCGCTT 5580

GTCTCTCTGG CAGCTTCATC AACTAATTTA TCGTATTCTT TCAAACCAAC TTGAACTACT 5640

GAAGGGCTAT TTGGATTATC AAATCCTAAA TATGTTTTTG TAGTTTCACT GCTAGTTGTT 5700

TTTAAAATAT CCAGGTAAGT AGATGGGTCT TGATAGTCTG GCCCCCATGA AACTCCTCCT 5760

GATACATCCC AATCCTCAGA TGAAGCATTG GCAGCATAGT AAGTAATATT AAGGAATTCA 5820

TCACTTGTCA TTTGTTGAAT ATCAACAACG ACATTTTCAA CACCAAGAAC TGTTTCTACA 5880

GATTGTTTAA AGGACTGAAT ACGAGATATG TAGTTTTTTG ATGCTTGGTC TACTGGAACG 5940

TCCAGATGAA TAGGAAACTG AACGCCGTCT GCTTCTAAAG CTTTCTTAGC TTTCGCAAAC 6000

TCTGCCTTGG CCTTGTCAGC ATTGAATAAA CCATCCTGCC CATCAGCTAA ATTCACACCT 6060

TTCCACTCAT CACCATAAGC AGGAAGTTGA GCAGCGACTA AATCACCAAA GGTCTTCTCA 6120

CCAGCTGAAA CAAAGTCTGG TTTTACAAAT AAATTACGAA CTGCTAAAGC TGCTCCATCT 6180

TTACCATTGA TTTGAGCTGA GTAAGCTGAG CGATCAAGAG CAAAATTCAA GGCTTGACGG 6240

AAATCTTTGT TAAGCAATGC CTTCTTAGTA GCTACTTTCT CTGAATCTGT AGTTTTAGAA 6300

GTATAGTTGT AACTTTGGCG ATCAATATTC ACACCCAGAC CAGCAATCCC AGAGCCTGAT 6360

TGTGTGTAAT AGATATTGTC CTTGTATTCT TCTGCAACCT TAGAATAGTT GGAGCTGGTA 6420

GGGTAAAGAC GGGCATAACT ATAAGCTCCA CTAGTGAAGT TACGCTCTAG CGACTCCTGA 6480

TCTGATCCAT CATAGTAAGC TAGATTGATA GTATCTAGGT GGACATTTTC TTTATCCCAA 6540

TATTGCTCAT TTTTTACAAA CTCTACAGAA GATTTTGCAG TCAACCCTTT CAACAAGAAT 6600

GGACCATTAT AAAGCAAGGA TGTCGGATCT GTTGGTTTAG CAAAATCGCT TCCTTTTGAT 6660

GTTTCGAATT CTTCATTCAG AGGCCAGAAA ATAGAATAGG TCAACTTAGA GTTCCAGAAC 6720

GGTTCAGGCT GGTTCAAAGT GTATTGTAAC GTATAATCAT CAACCGCCTT GACACCAACT 6780

GTTGAAAAAT CTGTTGAAGT TCCTGATAGA TAATCTGCCA AGCCTTTAAC CGAATTTTCA 6840

GCTAAATACA TAGCTTCTGA TTTTTTATCT GCTGCGTGTT TTAAACCGTT CACGAAATCT 6900

TTAGCCGTCA CCTCTGCATA TTCTTCTCCA TCAGAGGTAA ACCATTTAAC CCCTTTACGA 6960

ATCTTATAAG TGTAGGTCAA ACCATCCTTA GAGACTTCCC AATCCTCTGC AACTGCAGGA 7020

GCAAGATTAC CGTAATTATC GTTAGTGAAT AAACCATCAA TCCCATTTGA AGTCACTACT 7080

GTTGTACTAT TTTTACTTGA AATCAGGTAG TCCAAGGTTT CTGGGTCTGC TGTATAAACA 7140

TAGCCATAAG CTTTAGGGGC TGATGAATCA GATGATTTTG AAGAACTGCA TGCTGCAAGT 7200

ACACCTGCTG CTAATAAAAC AAGACCTGCT GTAGCAAATA CACGATTTTT TTTCATTTTC 7260

TACTCCTCTG TTTATGTGAA TTATAGATTG ACAACCATTA TATCACATTA TCCATTAAAA 7320

ATCAAACAAA TTTTCAGAAT ATTTAGGCTT GTTGGCACAA ATTTTTCATT TTTTTTGAAT 7380

ATATGATTCA AATTGTCGTT CGAAGTGTCA AAGACTACAG TGAAAATAGG AAATTTGACG 7440

CAGAAACTTT GGAGTTTAGG AAGACATACA GTAAAATGAA ATACGGACGG AACAATGTGA 7500

TTTTGGAATT CAAATTAAAT TATAACAATA TTGTAGAAGT ATCATTCTAG TATTCAAGAT 7560

TCAGTTTACT ATGTCTTTTC ACACCAACCT TATCCCGAAT TCAATTACTT TTGTGATTTA 7620

CATATATAGA TTAAGACTAT CTTTTATACT TTAAAATTTC TCGCTACCTT ATCCACTATA 7680

TGCTCCTCGC TATCACGTTT CTATTCATAG CCTACGATTT CACTATTGCT TTCTCTGACA 7740

ATTCTTATTT CCTGCGTCAG ACTTAAAACG ATCTATCCCC AGACCATTTT AATCCGCTAC 7800

CTCACGATAG TCAGGCTTGG GGAGCGCTAT TGTATTCACC GGTAGTGGAG CCCTACAGAG 7860

GACTTACACC TCAGATGCAC GACATGCCCA TCGTATAAAA AATCTCCTAC CCAAGGTAGA 7920

AGATTTCAAA CTTATAAAAC TTAATCCGTC ATGTCCGATA CCAACATTCG ATGCTCCAAT 7980

GGAATACTGC ACATAACTAG CAAGAAAATA AAGCCTGACT GAATCCAGAA GAGAGCCAAG 8040

TCAAAAATTC CGTGCACAGC AACCACTGTA AGGAAAGATA GATAAAGGCC GATAATCGGA 8100

CGTTTCCCCG ACTCCTGACT CATATCCATC ATCAAGCGAA CAGGAGCAAC AGAAGACAAA 8160

ACTAATAAAA TAGTCCCCAC AATTCCGTAA CTCAGAATCG TATCAATATA AAGACTGTGG 8220

GCATGTTCAT GATAAGGAGC ATGTATCCGA GGATAAGAGT TCATATAGGT CAATGGCCCT 8280

TCACCCCAAA AAGGATTTTG CTTAAACAAG GCCATCCCAG CATCCCAGAT AGAAATGCGT 8340

TCTTCCATAG AAGAGTCTAA AGTACCCATT CGAACTCCCA AATCACTAGA AAAGAGGAAA 8400

CTCAAACCAA TCGCGAAGAC CCCAATACTA AGCCAAAAGG CCTTCCAGTT TTTAATAGTC 8460

GTAAAGAGAT AGATAATTGC TCCAGCGATA ATAGCAGGAA AGGCAGTTCG ATTTTGAGTA 8520

AAGTTCAAAC CAAAGAGATT AACAAAGCCT GCAATCACAC AGAATACTTT CAACCAATTC 8580

AACTTGGTCG TTGTAAACAG ATAGAAAGCA ATCATAATAC AGAAACAACA AATAATTCCA 8640

TAATAATTAG GATTAAAGAA GGTCACTTCT GCCCGGTTCT GATGCCACAC CTGCATATTG 8700

GGTGAAAGAA AAGCATAGTT AAATTTCTTC ACAATTTGGA AATGTTCTAA ACTGGCAAAA 8760

GCAGCTGACA AGACACTACC AAACAAGACA AACTGCAAAA TCAATCGAAA GAATTTATGG 8820

GATAAAATCG ACTGATAGTG CAAAAAGAAA ATAGTAAATA GAAACATTCC TACTGAAGCC 8880

ACAAGACCCA TCCAATTTTG TGCAAGAATG GATATAACAG TACTATAGCT AAGAAAAAGA 8940

AGCAGCATCG GATGCTCCCC CATTTTCTGA AGAATACTTT TCATGTCTCC TGTAAAAATC 9000

AAACTGATAA TATATAAACA GAGTACAACT ACAAAAAGAT AAAAGGGTAA AAAGATACTC 9060

AGGATAATTC CCAATAAAAT CAGCTCTTTA CTAGACAACC CCTTCAGCTT TTCAATAAAG 9120

CCTATTGATT TCAAAATGAA TCCTTTCTCT CCAAATCAGC TGATTCAGAT AATAGTAAGC 9180

TATCCTATAT TGTACCACTT TTTTAGCAAT TTGAAAACAA AGGAAACGTT TTCCAAAATA 9240

AAAACCCTAT TTTATCCACC ATATCAAGGC TTCAAAATGA TACTTCAACT CCATTCTCAA 9300

TTACCCGATA AGTCTGATTT TGCAAATCAA TTTCTACTAC TGCTGTTACG GACTTATCTT 9360

TATTTTGACG TTTGATTACA ATGCTGTGAG CTGTTGGTGT CTCTATCTCA GTAGTCCCTT 9420

CTAGATCAAA GGCTTCTGAA CGGTTACGGA AAGAAAATAG ATTGAGAAGG GCCTTCACAA 9480

CAGGTCGTTG CACTTCTTTT GCTATTTCCT CGTTGCTATA GTAATGACGA TTAATATTTC 9540

GACCTTCTTT AGTTTCTTCT AATAATTTCA AGTCATTCTT GCCTGCTAAT AGACCCACAT 9600

AGTAAATCTG AGGAATACCT GGGGCAAAAG CTTGAATTAG ACGAGCGAGA AAATACTTGA 9660

CATCATCATC TCCAAGCGCT GAATAGTAGG TTGAATTGAT TTGGTAGATA TCTAAGTTGT 9720

TATACTCGGC ACTAGAGTAC TTACGTTTGA CATTGGCTCC AACCTTATAG AGTTCATTTG 9780

AAGCATAGTC AATCTCCTCA TCGGTCAGGA TATCCTTGAC ATCTACTACT CCAATCCCAT 9840

CATGGGTATC TAGCGTCGTA AATTGCTTCA TCGGGCTCAT CTTTAACCAC TTAGCCAAAC 9900

GCTCTGTTCT GGAACTGTAA AGAGTATAAA GTGTCACCAT TGGAAGAGCA AAATCATAAA 9960

CATAGTAATC ATGGTCTGCT ATTTTAAACT GAATCGAATA GTGTTCATGA ATCTCAGGTA 10020

AAAGCTCTGT CCCATACTCA GCAGCGATAT CTCGAACTTT GTCCAATAAA TCCCAAATAT 10080

CTGGTTCCAC AAAGAAATCA TTAGTATCCA ATTTCTTCAC TGCATAAGCA AAGGCATCTA 10140

GACGAATCAA ATCACACCCA TTACTTGCCA AGTGCTGAAT GGTCTTACGG ATAAATTCCA 10200

TAGTTACTTC TTTGGTCACA TCAAGATCAA TCTGCTCCTC ACCAAAGGTA TTCCACAAAT 10260

GTTCCACTGA ACCATCTTCA AACACAATCT CTTGCTTTGG TGCACGATCC TTACGCTTGT 10320

AAATTAAATC TACATCAGAC TGTGTCGGAC GGTTTTCTGG CCAAAACTTA TCCCAGTTTA 10380

AAAAGAGAGC TTTAAATTCA CTGGCTTCAT GTTTTTCTTG ATAGTCCTTA TAATACTTGG 10440

ATTGACGAGA AATATGATTA ATCATAAAAT CAAACATAAG ATAATATTTC TCACCTAAAC 10500

GCTTCACATC CTCCCAATCA CCAAAAGCTG AGTCCACTTC GTCGTAGTCA ACTGGCGCAA 10560

ATCCACGATC AACTGTTGAT GGGAAAAATG GTAAAAGGTG AACTCCTCCA ATAGCATCTC 10620

CAAAATGCTC TTCCAAATTA TCATATAAGT CTTTAAGATT ATTTCCAAGG CTATCAGAAT 10680

AGGTAATCAA CATGGTTTTA TTTTGAATTG GCATCATTAC TCTCCTTTTT CTAATTGAAG 10740

CCAAGTCTCA TATGATCTGG CTTCATAAAT AAAATTCATT TTAAATCTCT ATTTATCATC 10800

AAACTCGTAC TAATATAGAC TGTGATAAAC AAAGTACTAC TTTCTTGTTT TCTGCATAGA 10860

ATTATCAACA AGCTAAACTC TTCCTCTGTG TCAAAGACTA TAGATTCCAT GAGCTCTTCT 10920

TATACTCTTC GAAAATCTCT TCAAACCACG TCAGCTTCAC CTTGCCGTAG GTATGGTTAC 10980

TGACTTCGTC AGTTTCATCC ACAACCTCAA AACAGTGTTT TGAGCAACCT GCGGCTAGCT 11040

TCCTAGTTTG CTCTTTGATT TTCATTGAGT ATTACTTCAC TGCCCCGTTG CTCATTCCTG 11100

AAATGATATG GCGTTGGAAG AAGAGATAGA CAATGGTGAT ACTGATAATG CCGACCACGT 11160

AAGAGGCAAA GCTTGGTCCG TAGTCGTTGA AATATTGGCC TGCGTAGTTG TATTGGAACA 11220

AAGGCAGAGT CCACATTTTG GAATCCCGGT TCAAGACAAG GAGTGGCAAC ATGAAGTCAT 11280

TCCAGAACCA AAGGGCATTG ATGATCATGG TTGTCGCATG CATCGGTTTC ATCATTGGGA 11340

AGATGATGCG GAAATAGGTT GTAAATTGAT TAGCCCCATC GATCTCTGCT GCTTCATCCA 11400

GACTTTCTGG AATCGAGATT TTGATATAGC CAACATAGAG AAAGAGGGTC TGTGGAATCG 11460

CATAGGTCAA GTAGAGCAAG ATCAAACCAA AGGTATTAGC CAAACCGAGT TTACTCATCA 11520

TAACCGTAAT CGGAATCATG ATGACTTGGA AAGGTACGAA GATTCCGAGG ATTAAGAGGG 11580

TATACATGAT GGTAAAGGCT TTTCTTTTAC TCATATTGCG AGCGATGGAG TAGGCTGCCA 11640

TAGGGATAAA GATCATTACT GCAAGTAAAG ACAAGACAGT GATGACGACA GAGTTCCAAT 11700

AATAGCCTCC AATCCCATCA GCTAAGAGAC GGCTAAAGTT GTCCCATGTG AAGTTGGTTG 11760

GAAAGCCAAA GAAATTATCT ACAATATCCT TAGTGGGTTT GAAGGAACTA AAGAGGGTAG 11820

CAAGGAGCGG CACTAAAATC AGAACCGATC CTAGAATCAA TAGAATGTAT TTGCCAATCA 11880

GGGCTTTTCT TTCATCTTGT TTCATCATGC TTCTCCTCTT AAATTTCAAA TTTCTTAGAT 11940

ACTCTCAATT GGATGATCGA AATCACTACA ATTAAGAAGA ACAAGATTAC GGCAATGGCA 12000

TTGGCATAAC CGAATTGGTT GTTTTTAAAG GCATAGTTAT AAACCAAGAG CCCAAGTGAG 12060

GTTGTGGCAT TGTTTGGACC ACCACCGGTC ATGGCAAAGA CTTGGTCAAA GGCAGTCAGC 12120

CCACCTTTTA GGGCTAGGAT AAAGACCATA GAGACACTTG GTAGCAAGTA AGGCAATTCA 12180

ATGTTCCAGA AAACTTGCTT GCTAGTCGCA CCATCAATCC TTGCTGCCTC TGTAATCTCA 12240

GTTGGAATAG ATTGCAAACC AGCTAGGAAG ATGATGATGG GCATAGCCAC CCCTTGCCAA 12300

AGAAGGACAA AGACAGCCGC AAAGATTGCT CCCCACTTAG TCCCTAAAAG ACTGGTTTGG 12360

AAAAATTCAA TATGAAGGGC ATTTCCAATC GCTGGAAGAC CGTAGTTGAA GACTTGCTTG 12420

AAGATCAAAG CCACTGTCAA ACCAGATAAA ACAGCTGGGA AGAAGAACCA AGCACGGAAG 12480

AAGGTTTGGC CTTTGATTTT AGAATTCAAG ACACGCGCAA TGAAGATCCC GAGTGCAATC 12540

TCACCAACCA CCATGGCAAT CGCAATGATT GCGGTAAAGC CAATCGCATT CATGAATTTT 12600

GGATCCATGA AGAGGAGCTT AAAGTTGTTT AAGCCAACAA ATTTGTAGTT ATAAGTCAAT 12660

CCTGTCCAGT TGGTAAAACT GTAAAAGGCT CCTTGAAACA TCGGCACATA GAAGAAAATT 12720

GCTTGTAACA AGAGGGGGAT GACCACAAAA GCCCATGCCC AATATTTTTG TAATACTTTT 12780

TTCATAGTCT CTCTACTCCT AATCCACATC CGCTTTCATC GGGTTAAAGA AGGCATTCAA 12840

ATCATTGACC ATGCCTTGTT TATCACCGGT CAAGACATAG TTCATGGTCA AGGTATGGAA 12900

GTCTGCTTCA CTGGTCCAGT ATTGTTGCAA CCAGACCAAG TGACGATCCG TAAAGGCATA 12960

TTCGGTCATA CCAGCAAGCG GTGAATCTTC TCCTGCTTGT TTGACCCCTT CGATCGCTGT 13020

TGGAGATCCG TCCACATCGT AGTATTTTTG CATGACTTCT GGACGGGTCA TATATTCCAC 13080

AAAGGCATTG GCTTCTTTTG GATGTTTGGT GGTGGCTGAG ATAGACCATG CCAAGTCTCC 13140

CGCACCAACG GTTAAGCTTT GTCCTTTTTC TTTTCCTGGA ATCATGAAGG TCCCAATCTT 13200

AAAGTTCGGT TTTTGTTCAT TAATCGCTGT GATCGCCCAA GACCCATTTG GTGTCATGAG 13260

GACATCCCCA CGTGCGAAGG CTCCGATAAC ATCGGTATAG CCAGCACCTT CCCAGTTCTT 13320

TTGCTTAGAT CCATTGATGC GAAGGATGTC CATGACCTTG ATATCATCTT TCATAATCGG 13380

ATCCGACAAT TTAATGGCAT TTGGTTGAGA ATAACGAAGG TATTGATTTG CTTCTTTTCC 13440

TCCACCTGTT GCTGTCGCAA AGGCTAATTG ATTGTAACCA TTGAGTGTCC AAGCATCTGC 13500

ACCTGCAATT CCAAATGGTG TTTGTCCTTT AGCAACGATA TCTTTGACTA ACTGTTCAAA 13560

TTCATCCCAG GTTTCAGGAA CCTTCAAGCC CAGTTCTTCG AATTTATCTT TGTTGTAGTA 13620

AATTCCATAA GCATTAGCTG TAAAAGGAAC GTTGTAAACT TTTTCGTTTA CAGCATATTT 13680

TTCAGCGTAG CCATTTTTCA CGCGTTTCAG GTAGTCTTTG TTGCTCAAAT CTTCAAAAAC 13740

ACCTGCTTTT GCCCATTCTT GCAGTTCGAT GGACTGTGGG TAAATATTGA CCACATCAGG 13800

CACATCTCCT GCGAGAACGC GTGTCTTCAA TACTTCACCA GCATTTGGTA CATTGACGAC 13860

TTTGACCTTG ATCTTAGGGT TTTCCTTCTC AAAATCACGA GTGATTTCTT CCAAGGTTTT 13920

GGTCATTTCT TTTTTCTGGT TGAAATACTC GATGGTCACT GTGCCATCCG CAGATTTACC 13980

ATAGTTGGAG CAAGCGCCGA GCCCAAACAA AGCTAAACCT GTAGTTGCAA GAAGTCCGAT 14040

TTTTTTATAC CATTCCATTA GAAAGCCTCC TTTATAAATT TATACACCCT TATTGAACTG 14100

CACCCCAAAA GTTAGACAGA ATAAATCTAA CTTTTGGGGT CAGTACATAT CATAGTTTTC 14160

TAAAAATATA CTGTCTACTC AAAAAATCTC CTTGGGATAA GATAACAGTT AAGCCCGCAT 14220

ACATTAGTTC TGCACCTGAG TAAACTTCGC CATTTTCCTG TAATTTATAT AGTCCCTCTT 14280

CATCCAAATC TTTTAATTTT AAAGTTGTTT CCATGGTCTC TACAACAGAT AAAACGCGAA 14340

CGTAGGTTAC AATCGTTTGA TTTCCGTAAT TAAATTGTAC AGCTGCTTCA TTGGATACAG 14400

TATCAGGATT AATTAGTCTA TACTGCTGTC CTAACTGAAC TACTGGTCGT AATTCTTTAT 14460

ACAAGTTCAC CTGATTAGCA ATCGTAGCTT TCTCTTCATC TGATAAATTT GTCAAATCAA 14520

GTTCATAGCC CAAATTTCCC ATCATTGCTA CAAGGCCACG TGTTTCTAAT GGTGTCATTC 14580

GTCCCATCTG ATGATTCGGT ACTGCTGACA CATGAGCCCC CATAGAAATG GTTGGATAGA 14640

GATAGGATGA ACCGTATTGA ATTGGTAAAC GTGCAATGGC ATCAGTATTA TCACTAGCCC 14700

AGACTTGTGG GAAATAGCGC ATCATACCAA GATCATTTCG TCCACCACCA CCAGAGCAGG 14760

ACTCAAAGAG AATATGGCTG TGCTTCTCTG TCAGATAAGA AACGAGTTCA TAAAGCCCCA 14820

GCATGTACTG ATGAGATTGC ATCTGTGTCT CTAGATAAGT TAATCCATTC CCTAGCTTAG 14880

TGATATTGCG GTTCATATCC CATTTAATGT AATCAATATC ATGATAAAAT AGGAGTTGAT 14940

CTAAGACACT TTTCAAGTAT TCTACTACCT GAGGATTGGC AAGATTAAGT ACTAATTGAT 15000

TCCGAGAATA AGTATGCTCA TAGCCAGGAA CCTGAATAGC CCAGTCAGGA TGTTGACGAT 15060

ACAAATCACT ATCTACAGAA ATCATTTCGG GTTCTAACCA AAGTCCAAAC TGCAAACCTC 15120

TTTCATGGAT AGCTGAAATC AGACTTTCTA GACTTCCACC CAGTTTTTCC TCATTAACAA 15180

CCCAATCACC TAAAGCACGA TTATCATCAA AACGATTGCC AAACCAACCA TCATCTAATA 15240

CAAAAAGTTC AATGCCAACT TTCTTAGCTT CATCTGCTAA CTCTAACAGT TTTTCTCTCT 15300

GAAAGTCAAA GTAAGTAGCT TCCCAGTTAT TGATTAGAAT TGGACGTTCT TTTTTAGAAA 15360

ATTCACTTAG CATAATGTGC TTCAGTACAA AATTCTGACT TTCATGACTA ATACCAGTTA 15420

ATCCCTGATC TGAATGAGTC ACTAAAGCTA CCGGTGTTTC AAAGTATTCC TCAGGAGCTA 15480

ACTTCCAAGA AAAGTTTTCT GGATTAATGC CAATAGCCAC CCGAACTTCA TTCAATTGAT 15540

TTTTTTGAAC AAAAGCTTCA AAGTTGCCAC TATACATTAG TTGAATAGCA AACACATTCC 15600

CAGCATCCTC TGTGACTCCT TGTTCGCATA GTAGAAGAGC TGGTGTTTGA GCATGACCAG 15660

AAGCACCTCG GTTTGAACTA ATCGAAAAGA TTCCTTGTTC TACCTGTTGA CGTCTAACAG 15720

TCTTTTCACG AGCATAAGCA CCCTGCAGAG TTACTATTTC GTAATCTGCA GCTGGAAAAT 15780

CAGCCATAAA AGAAAAATCT TTATGGATGA CAACTTCCTG ATTACTATTA TTATCTAATT 15840

TACTGTAGCT AGCAATAGTC GCATCATTAT TAAAAGTAGT ATAATACAAA GTCAGACTAA 15900

GTTGAGCCTT AGAATCTTCT AACATTAAGA CAAGAGTCTC TGTATCGTCC ATGCTATGTG 15960

GAGAAGGTAA GCCCTGTGGA CCATTCTGAC CTTTTAAAAT CTTTGCTTCT ACAAATCGAA 16020

AGTCTGTTAC TTCAGTTACA CTATGCTGAA CCTGTATGGT TGGTTTCCTA AAATCTCCTA 16080

AGCCATGTTG TCCAAAAATC TGTCGCTGAG TATCTAAACT AAAGGTTCGA TTAGTAGCCG 16140

TTGGATTTCC TGAAAAGGCA TGGTCTCGTT CATAAACACT ATTGGAACCT TTATAGTTCT 16200

TAATAGTCTT TCCTAAATGT TTCAAAAGTA AGTAGCCATT TCGATTTTCA ATAATCAAAC 16260

TTAGATTTTT ACTCTCAACA TAAAATAGAT TATTCTCTAT CCTAACTCCC ATTTACTTCA 16320

CCTCATCACT TTATTGATTA TATTTTATCA CCTGAAATCG CTTTCCAAAA TAGAAAAATG 16380

TCTCAAGAAT ATGGTAAAAT GTTAGGTAGG AGGTAGCACA TGTTAGTTTT TTCAGAATAC 16440

CAGACTGGAA CAATCGACCT TGCCCTAAGC TTTTATGGAT ATGAGGAATG CACACCTAAT 16500

TACTCTTTTG GTCCAGCCAT TCGTGATACA TACGTTCTAC ATTACATTAC TAAAGGACAA 16560

GGAAAATTTC ATTACAAGGG TAAAATTGTT GATTTAAAAG AAGGAGATTT CTTTCTATTA 16620

AAACCAGAGG AACTAACCTT TTATCAAGCA GATAGTAAAG AACCTTGGGC CTACTACTGG 16680

TTAGGAATCA CTGGAGGGAA AGCCCCTGAT TATTTTGCTC TTTCCCAAAT TTCTGATCAA 16740

TCCTATCTCA TCCAATCTGA AACTTGTCAT ACCCAGACTA CTGCAAAACT CATCTCAGAC 16800

ATTGTCCGCT TCGCTCAGAT TACAAAATCA AGTGAATTAG CTCAACTCCA TATCATGGGA 16860

CAACTTCATG AACTGATGTT TCATCTGGGA ACTATTGCTC CCAATCAGAA AAAAAAGAAT 16920

ATTTCATCAA CCCACCAACT CTATCTTGAA TGCAAACGAT TAATTGATAG CCACTATCCT 16980

CAATCACTTA CAATTCAAGA TTTAGCAAAA GAACTATCCG TTCACAGAAG CTACTTATCA 17040

AGCGTATTCA AAGAATTTAA TACCTTATCA CCCAAAGAAT ACCTACTCTA CGTTCGAATG 17100

CACCGAGCTA GACAACTTCT CGAAAATACC CAAGAGTCCA TCAAGGTAAT TGCATACTCG 17160

GTAGGTTTTT CAGATCCACT CCATTTTTCG AAAGCTTATA AACAATACTT TAATCAGACT 17220

CCAAGTCATA CAAGAAAAGA ATACTCTCAA TACCAACTAG TAAGAAAGGC AACATTATGA 17280

AATCCTACCA AGCTGTCTAC CAAATCCTAT CTAAAGAAAC CGACTATATC AGCGGAGAAA 17340

AAATCGCAGA AAAACTATCC CTAAGCCGAA CAGCAATTTG GAAAGCCATC AAGCGACTAG 17400

AACAAGAAGG CATTGAAATT GATAGTATCA AAAATAGAGG ATATAAACTG ATGAATGGTG 17460

ACCTTATTCT TCCAGAGATT CTAGAAGAAA ATCTTCCAAT TAAAGTCAGC TTTAAACCCG 17520

AAACAAAATC AACACAACTA GATGCAAAAG AAGCAATTGA TTTAGGCCAT GAAGCAAATA 17580

CCCTCTATCT AGCTTCCTAT CAAACAGCAG GCCGAGGCCG TTTTCAACGT TCCTTCTACT 17640

CACCACAAGG TGGTATTTAT ATGACACTCC ATCTTAAACC AAATCTCCCC TATGACAAAT 17700

TACCATCCTA CACACTACTT GTAGCTGGAG CTGTCTACAA AGCCATTAAG AACCTAACTT 17760

TAATAGATGT CGACATAAAA TGGGTCAATG ATATCTATCT AAACAATCAT AAAATTGGAG 17820

GAATCCTTAC TGAAGCAATG ACCTCTGTAG AAACTGGCTT AGTCACAGAT ATCATTATTG 17880

GAGTAGGTAT CAATTTCACT ATTAAAGACT TCCCTCAGGA ATTAAAAAGAA AAAGCTGCCA 17940

GCTTATTTAA AGCTACAGCT CCTATAACAA GGAATGAATT GATCATAGAA ATCTGGCGTG 18000

CTTTCTTCGA AACACCAGCA GAAGAGCTAT TATACCTATA CAAAAAACAG TCATTCATTC 18060

TAGGAAAAGA AGTCACTTTC ACACTAGAGC AAAAAGACTA CAAGGGACTT GCTAAAGACA 18120

TCTCAGAAAA TGGAAAACTT TTAGTTCAAT GTGATAACGG AAAAGAAATC TGGCTAAATA 18180

GTGGCGAAAT TTCTCTCAAT AGTTGGAAGT AAAATAACAC AATTATAATA TAAACGATAT 18240

AAAAATAACT TCAGATTAGT AATTCAATTA AGTTTTACGG ATCTGAAGTT TTATTGGCTC 18300

TAAAAATAAA AAAGAGAGTT ACAGACTCTC ATTAAAACGG AGAATAAGGG ATTCGAACCC 18360

TTGCGCCAGT TACCCGACCT AACGATTTAG CAAACCGTCC TCTTCAGCCT CTTGAGTAAT 18420

TCTCCAATTA ATGGGCACGA GTGGACTCGA ACCACCGACC TCACGCTTAT CAGGCGTGCG 18480

CTCTAACCAC CTGAGCTACG CGCCCAAGTT AAAAAACTTG GTAATTTGAA CAAAGTTCAA 18540

AGCGGGTGAC GAGAATCGAA CTCGCGACAA CAGCTTGGAA GGCTGTAGTT TTACCACTAA 18600

ACTACACCCG CATAAATACT ATCAATAAAA TGGCGCGAGA CGGAATCGAA CCGCCGACAC 18660

ATGGAGCTTC AATCCATTGC TCTACCAACT GAGCTACCGA GCCTTATTGC GGGAGCAGGA 18720

TTTGAACCTA CGACCTTCGG GTTATGAGCC CGACGAGCTA CCGAGCTGCT CCATCCCGCG 18780

TTAATAATAT AAAAGGAGGA TGTGGGATTC GAACCCACGC ACGCTTTTAC ACGCCTGACG 18840

GTTTTCAAGA CCGTTCCCTT CAGCCGGACT TGGGTAATCC TCCAATATTC AAATGGACCT 18900

TGTAGGACTT GAACCTACGA CCACTCGGTT ATGAGCCGAG AGCTCTAACC AGCTGAGCTA 18960

AAGGTCCGAC AAGATCATTA TAGCGGCGAA GGGGATCGAA CCCCCGACCT CCCGGGTATG 19020

AACCGGACGC TCTAGCCAGC TGAGCTACAC CGCCATGAAT CGGGAAGACA GGATTCGAAC 19080

CTGCGACACC TTGGTCCCAA ACCAAGTACT CTACCAAGCT GAGCTACTTC CCGAGTTAAA 19140

TAGAAAAATG CACCCTAGAG GAGTCGAACC TCTAACCGCC TGATTCGTAG TCAGGTACTC 19200

TATCCAGTTG AGCTAAGGGT GCTCCATATT ATGCCGAGGA CCGGAATCGA ACCGGTACGA 19260

TCGTTACCAA TCGCAGGATT TTAAGTCCTG TGCGTCTGCC AGTTCCGCCA CCCCGGCCTC 19320

TCTAAGCGAA CGACGGGATT CGAACCCGCG ACCCCCACCT TGGCAAGGTG GTGTTCTACC 19380

ACTGAACTAC GTTCGCACTG TTTTCTTCTA TCTAAAAATG CCGGCTACAT GACTTGAACA 19440

CGCGACCCTC TGATTACAAA TCAGATGCTC TACCAACTGA GCTAAGCCGG CTCATTTGTT 19500

ATATCTTAAT GCGGGTTAAG GGACTTGAAC CCCCACGCCG TTAAGCGCCA GATCCTAAAT 19560

CTGGTGCGTC TGCCAATTCC GCCAAACCCG CATATATGAC CCGTACTGGG CTCGAACCAG 19620

TGACCCATTG ATTAAAAGTC AATTGCTCTA CCAACTGAGC TAACGAGTCT AAAATAACTT 19680

GCGTTACCTT AAACGGTCCG ACGGAATCGA CCCGGTAC 19718

(2) INFORMATION FOR SEQ ID NO: 100:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4117 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

CCGTGGAAAA GTCTGGATAG TGAATGGTCT TCACACAATG ACCTGAAAGA AGCCTGAGAA 60

TAATTATGGA GAGTAGCATT CTGAGAGGTG TTAGCAGAAC CATATGACAG AGCTGTTTGA 120

AGAGGGAATA TTGAGGAGAA AAATCCTGAG CCTACCAGTT GGAGTTGGAA AGAGCTGACT 180

GTTAGATCAT GGTTTATTAT CCACAACCTG TGGATAACTT TGTGAATAAG AGAAGTTGCT 240

AAAGAAGGAG ATATATAACG ATGAAGAAAA TCAAACCGCA TGGACCGTTA CCAAGTCAGA 300

CTCAGCTAGC TTATCTGGGA GATGAACTAG CAGCTTTTAT CCACTTCGGT CCTAATACCT 360

TTTATGACCA AGAATGGGGG ACTGGACAGG AGGATCCTGA GCGCTTTAAC CCGAGTCAGT 420

TGGATGCGCG TGAGTGGGTT CGTGTGCTCA AGGAAACGGG CTTCAAAAAG TTGATTTTGG 480

TGGTCAAGCA CCACGATGGC TTTGTCCTTT ATCCGACAGC TCACACAGAT TATTCGGTTA 540

AGGTCAGTCC TTGGAGGAGA GGAAAGGGCG ACTTGCTCCT TGAAGTATCC CAAGCTGCCA 600

CAGAGTTTGA TATGGATATG GGGGTCTACC TGTCACCGTG GGATGCCCAT AGTCCCCTCT 660

ATCATGTGGA CCGAGAAGCG GACTACAATG CCTATTATCT GGCTCAGTTG AAGGAAATCT 720

TATCAAATCC TAACTATGGG AATGCTGGTA AGTTCGCTGA GGTTTGGATG GATGGTGCCA 780

GAGGAGAGGG CGCGCAAAAG GTTAATTATG AATTTGAAAA ATGGTTTGAA ACCATTCGTG 840

ACCTGCAGGG CGATTGCTTG ATTTTTTCAA CAGAAGGCAC CAGTATCCGC TGGATTGGCA 900

ATGAACGAGG GTATGCAGGT GATCCACTGT GGCAAAAGGT GAATCCTGAT AAACTAGGAA 960

CAGAAGCAGA GCTGAACTAT CTTCAGCACG GGGATCCCTC GGGCACGATT TTTTCAATCG 1020

GAGAGGCAGA TGTTTCCATC CGTCCAGGCT GGTTCTACCA TGAGGATCAG GATCCTAAGT 1080

CTCTCGAGGA GTTGGTCGAA ATCTACTTTC ACTCAGTAGG GCGAGGAACT CCACTCTTGC 1140

TTAATATTCC GCCGAATCAA GCTGGGCTCT TTGATGCAAA GGATATTGAA CGACTTTATG 1200

AATTTGCGAC CTATCGCAAT GAGCTCTATA AAGAAGATTT GGCTCTGGGA GCTGAGGTAT 1260

CTGGTCCAGC TCTTTCCGCA GACTTTGCTT GTCGCCATTT GACAGACGGC CTTGAGACCA 1320

GCTCTTGGGC AAGCGATGCA GACTTGCCCA TCCAGTTAGA ACTCGACTTA GGTTCTCCTA 1380

AAACTTTTGA TGTAATTGAG TTAAGAGAAG ATTTGAAGCT AGGGCAACGA ATCGCTGCTT 1440

TTCATGTGCA AGTAGAGGTG GATGGTGTCT GGCAGGAGTT TGGTTCGGGT CATACTGTTG 1500

GTTACAAACG TCTCTTACGA GGAGCAGTTG TTGAGGCACA GAAGATACGT GTAGTCATTA 1560

CAGAATCACA GGCTTTGCCT TTGTTGACCA AGATTTCCCT TTATAAAACT CCTGGATTAT 1620

CAAAAAAAGA AGTTGTTCAG GAACTAGCAT TTGCAGAAAA AAGCCTAGCT GTGGCAAAGG 1680

GAGAAAATGC CTATTTTACA GTTAAGCGCA GAGAATGTAG TGGTCCTTTA GAAGCTAAGA 1740

TTTCGATTCA ACCGGGGACA GGTGTCCATG GTGTCGCCTA TCAGGATGAG ATTCAAGTCC 1800

TTGCGTTTCA AACTGGTGAG ACTGAAAAAA GTCTGACGCT ACCAACCTTG TATTTCGCAG 1860

GAGATAAAAC CTTGGATTTC TATCTGAACC TAACGGTGGA TGGTCAGCTT GTGGATCAAC 1920

TTCAAGTCCA AGTTTCATAA AAGAAGAACC TTTGCGCGAT GCAAAGGTTC TTTTGGTTAT 1980

TAGTGACTTG GTAACCAGCT GAGGGTGAAA GTTAGTTGTT CAGCTTTTAA GAGGTCTTGG 2040

TGTTGAATAG TTGATACGAG TGTTTTGTCC AGTCGGCATT CTTTGACAAA GTTAAAATGG 2100

TTGTGGTTTT GTTTAGTATG GATATCCAGC CATTTATCTT CTTTAGCGAG GTAGACTCGT 2160

AGATGGTCAA AGAGAGGGAT TCCGAGGTCA TAGCTTGGTT TTCCTGGACA GGTTGGATAA 2220

AATCCGAGAG CTGACCAGAT GTACCAAGCA GAGAGACTAC CATTGTCTTC ATCTCCAGGA 2280

TAGGCTTCCC AACTTGGGTG AAAAGCTTTC TGACGGAGCG TCTTGATAAG AAGGGCAGTG 2340

TAGTCAGGGT AATCGCTGTA ACGGAAGAGA TAAGGAATGT GGAAACTAGG CTGGTTGGAA 2400

ATGGCTATTT GTCCAAAAGG AGCAGTAGCC ATCTCGCTCA TTTCGTGAAT TTCGTAACCA 2460

TAGCCTGTTG TTTCAAAGAG GGGAGCATCT TGACAGGCTT TCAAAAGATA GTTGCTAAAG 2520

GTTTCTTTTC CACCCATCAG TTGGATTAAG CCAGGGATGT CGTGGAGAAC GCCTAAAGTA 2580

GCTTGAATGG CAGAGCATTC AGCGTAGTCT CGCCCCCAAC TATAAGGAGA GAAGTCAGGG 2640

TGAAAGTTTC CTTGATTGTC TCGTGCTCGC ATGTAACCTG TCTCAGCGTC AAATAGCTGG 2700

CGGTAATTTT GTGAAGCAGC CTTGTAGGTT TCAGCGATTT CTATGTTCTC TAGTTTTTTG 2760

GCACAGCTGG CGATACAAAA GTCACTATAG GCATAGTCTA GAGTATGGCT AACACTTTCG 2820

TGGTGGTCGG TAGAGAGGTA ACCTAGTTCT TGGTATTGGG CTAGTCCGTG GCGGCCATTG 2880

ATGCCGAGAG GGTCGGCTTT GCTGGCTGTT TCGAGCATGG CTTGGAAGAG TTCTCCTTCT 2940

AGGTCGGGGG TCATGTCCTT GCAGGCGCTA TCTGCGATAA TACCGTCTAA AAGTGTACCT 3000

GGCATCATAC CCCGTTCATC TGGAGCCAGC CATTTTGGAA GGAAACCAGT ATCGCGGTAG 3060

CTATTGAGGA AACCTTCTAA AAAGCGTTGA TAGTGCTCCG GTATGATAAG GGCAAAGAGG 3120

GGGAAGGTGG TGCGGAAGGT ATCCCAGAAA CCATTGTTGC TAAAGAGGAC ACCAGGCTTG 3180

ACAGTACCAG TAGCCAGATC CATGTGGATG GCTTGCCCTG ATTCATTAAT CTCATAAAAA 3240

GTCTGTGGGA AGAGGAAGAG TCTGTAGAGG CAGTGGTCAA AGAAGGTTCG GTCAGCCTCT 3300

CCTGTCTCTA TAATGTCAAA ACGATGGAGG AGATTTTCCC AATCCACTTG GGCACTTGAT 3360

TTACAGCTAT CAAAATCTTC TTGAGGTAGA TTGATTAGAG CTTGAGAAGG AGAGATGAAA 3420

GAAGTGGCTA GTTGCATCTC GGTTTGACTA CTTGCTAAGT CAATTCGCCA GTCTCCAGCT 3480

TCTTGGCTGA TAGCAAGAAT ATCCGTGTTC ATTTGCAGGG CAGTGAACAT CGTTAGCGAA 3540

TTTTTGTTAG TTTCAGTTTT ACCTTCTTGT CGCAGGGCAA GAGTCCGCTT ATCTACTTGC 3600

TCTACTGTCA GTTCATCTGC TGCGTGAAGA TAGAGGGAGA GGGCTTTGCC TTGCTTTTGA 3660

TTCAAACGAA TAGAAGCACC ATAGCAAGTC GGTGTGAGCT GGGTTTCAAT CTGATAACGC 3720

AGAGAAAAGA GCTTCAAATA GTGAGGCTGG AAGCAAGCTT TATCTATATC ATAAGAAGAC 3780

TGGCGGTGAA AGAGGCTGTC TCCCCCCAGT TGACTGGTGA CAGGTGTCAG AAGGAGCCAA 3840

GAGTAGTCCC CAATCCAAGG ACTGGGCTGG TGAGTTAATC GAATCCCCTG AAAGATAGGC 3900

AGATGTGGAT CAAAAAACCA AGATCCATCC TGGTCACTGG TCTGGGGCAC AAAGTAATTC 3960

ATCCCAAAAG GCACGCCTGT GTATGGCAGG GTATTTCCCC GAGAAAAGGC ATGCTTGTTG 4020

GTAGTTCCAA AACGGGTATC GATGGTATCA AGTAGTGGTT TCATAGTCTT TCCTTTAGCT 4080

GTTTTTCTAC ATTATATCAG TAATAGAGGG CCTTTAG 4117

(2) INFORMATION FOR SEQ ID NO: 101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2727 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

CTGGTTCAAT TATTATTCAC TCTAAGTAGT CATATGTTCT TTATTTATGT GAGTTTTTAC 60

CTTTTAAAGG ATCTTGTTAG ATGGGAGAAG GTTTTAAAAG TGACAGATGA TAATACAAGA 120

AAAGTTCGTT TATTAGTAGC CTTTTTTAGC ATTGTCATAG GCTACATCCT GAGTTCTTTC 180

TTTATTAGCC TGTATCATTT GTGGCAAGAA GCGCTTAGAG GATTATTATG AAATCAAGAG 240

TAAAGGAAAC GAGTATGGAT AAAATTGTGG TTCAAGGTGG CGATAATCGT CTGGTAGGAA 300

GCGTGACGAT CGAGGGAGCA AAAAATGCAG TCTTACCCTT GTTGGCAGCG ACTATTCTAG 360

CAAGTGAAGG AAAGACCGTC TTGCAGAATG TTCCGATTTT GTCGGATGTC TTTATTATGA 420

ATCAGGTAGT TGGTGGTTTG AATGCCAAGG TTGACTTTGA TGAGGAAGCT CATCTTGTCA 480

AGGTGGATGC TACTGGCGAC ATCACTGAGG AAGCCCCTTA CAAGTATGTC AGCAAGATGC 540

GCGCCTCCAT CGTTGTATTA GGGCCAATCC TTGCCCGTGT GGGTCATGCC AAGGTATCCA 600

TGCCAGGTGG TTGTACGATT GGTAGCCGTC CTATTGATCT TCATTTGAAA GGTCTGGAAG 660

CTATGGGGGT TAAGATTAGT CAGACAGCTG GTTACATCGA AGCCAAGGCA GAACGCTTGC 720

ATGGTGCTCA TATCTATATG GACTTTCCAA GTGTTGGTGC AACGCAGAAC TTGATGATGG 780

CAGCGACTCT GGCTGATGGG GTGACAGTGA TTGAGAATGC TGCGCGTGAG CCTGAGATTG 840

TTGACTTAGC CATTCTCCTT AATGAAATGG GAGCCAAGGT CAAAGGTGCT GGTACAGAGA 900

CTATAACCAT TACTGGTGTT GAGAAACTTC ATGGTACGAC TCACAATGTA GTCCAAGACC 960

GTATCGAAGC AGGAACCTTT ATGGTAGCTG CTGCCATGAC TGGTGGTGAT GTCTTGATTC 1020

GAGACGCTGT CTGGGAGCAC AACCGTCCCT TGATTGCCAA GTTACTTGAA ATGGGTGTTG 1080

AAGTAATTGA AGAAGACGAA GGAATTCGTG TTCGTTCTCA ACTAGAAAAT CTAAAAGCTG 1140

TTCATGTGAA AACCTTGCCC CACCCAGGAT TTCCAACAGA TATGCAGGCT CAATTTACAG 1200

CCTTGATGAC AGTTGCAAAA GGCGAATCAA CCATGGTGGA GACAGTTTTC GAAAATCGTT 1260

TCCAACACCT AGAAGAGATG CGCCGCATGG GCTTGCATTC TGAGATTATC CGTGATACAG 1320

CTCGTATTGT TGGTGGACAG CCTTTGCAGG GAGCAGAAGT TCTTTCAACT GACCTTCGTG 1380

CCAGTGCGGC CTTGATTTTG ACAGGTTTGG TAGCACAGGG AGAAACTGTG GTCGGTAAAT 1440

TGGTTCACTT GGATAGAGGT TACTACGGTT TCCATGAGAA GTTGGCGCAG CTAGGTGCTA 1500

AGATTCAGCG GATTGAGGCA AGTGATGAAG ATGAATAAGA AATCAAGCTA CGTAGTCAAG 1560

CGTTTACTTT TAGTCATCAT AGTACTGATT TTAGGTACTC TGGCTCTAGG AATCGGTTTA 1620

ATGGTAGGTT ATGGAATCTT GGGCAAGGGT CAAGATCCAT GGGCTATCCT GTCTCCAGCA 1680

AAATGGCAGG AATTGATTCA TAAATTTACA GGAAATTAGG CTGGAGAACC AGCCTTTTTC 1740

TAAAGATAAG GAGAAATATG AACAAAAAAA CAAGACAGAC ACTAATCGGA CTGCTAGTGT 1800

TATTGCTTTT GTCTACAGGG AGCTATTATA TCAAGCAGAT GCCGTCGGCA CCTAATAGTC 1860

CCAAAACCAA TCTTAGTCAG AAAAAACAAG CGTCTGAAGC TCCTAGTCAA GCATTGGCAG 1920

AGAGTGTCTT AACAGACGCA GTCAAGAGTC AAATAAAGGG GAGTCTGGAG TGGAATGGCT 1980

CAGGTGCTTT TATCGTCAAT GGTAATAAAA CAAATCTAGA TGCCAAGGTT TCAAGTAAGC 2040

CCTACGCTGA CAATAAAACA AAGACAGTGG GCAAGGAAAC TGTTCCAACC GTAGCTAATG 2100

CCCTCTTGTC TAAGGCCACT CGTCAGTACA AGAATCGTAA AGAAACTGGG AATGGTTCAA 2160

CTTCTTGGAC TCCTCCAGGT TGGCATCAGG TCAAGAATCT AAAGGGCTCT TATACCCATG 2220

CAGTCGATAG AGGTCATTTG TTAGGCTATG CCTTAATCGG TGGTTTGGAT GGTTTTGATG 2280

CCTCAACAAG CAATCCTAAA AACATTGCTG TTCAGACAGC CTGGGCAAAT CAGGCACAAG 2340

CCGAGTATTC GACTGGTCAA AACTACTATG AAAGCAAGGT GCGTAAAGCC TTGGACCAAA 2400

ACAAGCGTGT CCGTTACCGT GTAACCCTTT ACTACGCTTC AAACGAGGAT TTAGTTCCCT 2460

CAGCTTCACA GATTGAAGCC AAGTCTTCGG ATGGAGAATT GGAATTCAAT GTTCTAGTTC 2520

CCAATGTTCA AAAGGGACTT CAACTGGATT ACCGAACTGG AGAAGTAACT GTAACTCAGT 2580

AAAAGATACG CCTACACTCC TATGTCACTT ATGGATGTAG GAGTTCTTTT TACTAGTTTA 2640

AGCAGGACTA AGACAGGTAC TAAGACAAAA TAGCAACTTC TAAAACTAAC TTCCAGTTTT 2700

GGGAGAGAGA TGGAAGTTAC TTTGAGA 2727

(2) INFORMATION FOR SEQ ID NO: 102:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5717 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

TTTTTTGTAG ATTTAAGTGG GGTGCAATTC CTAAAAAATA AAAAACAATT TTTGAAAATT 60

ATGTTAGCAG GAATTGCTTC AAATTCGATT TTATCACTTA CAGGTTTACT TGTTTTATTG 120

TTCACATCGT ATAAATTGCT TGGACTCTTA TTTTTTATCA TTAACTTAGG TATGATTTTT 180

ATTAATTCAA TTCCTTTTTT TCAGTATGAT AGTGGTATTA TTTTAAGATA CTTGAATTCT 240

AACAATAATA ACTTGAATTT TCAATATATA GTTCAACTTT TAATAGCATT TGTTATTATT 300

TATTTTCCTT TGAGTCAACT ATTACAGTTT TTGACACCCA ATATTATTGT TCGTAGTATA 360

GGAGGGGTGG TTGTTTCTAT ACTGCTTTCT ATATTATATA TGATAGGAAG GACGAAATAT 420

GTTCTACGTA AATAGTTATG TTTTTGCTTA TAAAAAAGAA GGTATAATGT ATTTACGTGG 480

TCGGAGTATG CGGGAAATAG CTATAGAACC TCAAATTTCG CAAGAATTTA TCAACGATCT 540

ATTTAATAGT TGTAAGGAAC TATTAGAGAT AGAAGAAGTA TTAGGCAGTA AACTAACATT 600

TGAACTATAA ATGAACAAAT TTTAATTTCG GATGAGATAG ATATTGATAG TAGATATTCT 660

AGAACTAAAG GTTACTATTC GTTATTTTAT AATGAAGAGT ATAATAAAAT ACAGAATAAA 720

ACAGTATTAG TATTAGGAGC AGGAGTCTTA GGATGTTATA TATCTCTAAG TCTAAGTATG 780

TATGGAGTGA GGAAACTTAT TGTCGCTGAT TACGATATAA TAGAACCATC AAATTTAAAT 840

AGGCAAATTC TTTATACAGA GTCGGATGTT GGTAAGGAGA AGATTAATGT TCTTTCTGAA 900

AAAATACACA AGTATAATTC AGATGTTCAG GTAGTACCTA TTTCTATTAA AGTTTCTTCA 960

GTAGAAGAAT TAGAAAAAAT TGTTGCGGAA TATGGGAGTA TAGATTTTAT CGTTAAAGCA 1020

ATTGATACGC CCATTGATAT TATAAAAATT GTCAATCAAT TTGCTGTATC GCATAAGATA 1080

TCCTACATAT CAGGAGGGTT TAATGGATGC TATCTTATTA TTGATAATAT ATATATCCCT 1140

ACCATCGGTT CTTGCTTTGG TTGTCGGAAT ATAAACAAAG ATATAAATAA GTACACTTTA 1200

TCTGATAAGA CAAAGTGGCC GACTACACCA GAGATGCCTG CTATTTTGGG AGGGATAATG 1260

ACTAATTTAA TAATTAAAAT ATTTCTGGGA TGTTATAATG AAATCCTAAT AGATAACGCT 1320

TACGTTTATA ATATGAGAAA TCATGCTCTA AGTCAAGAAA AATATGTTCT GGAAAACGGA 1380

GAATGTCCAA TTTGTAAAAA AATAATAAAG TGAAAGATAA CAATATTAGA GCGAAAACAT 1440

TTATTCGTTC AGTTTGTTTT TGCTTATTAT CAGGAGGAGT AGCTTTTTTA TCTGCTATTG 1500

GGCAGTTCAC TGTTATAGAA ACACAATTAA TAGTATTGTT CTTGGGTATT ATTTTTGCTA 1560

TATATTATGC TTACTACAAT AAAAATATTC AAACATCATT GGAAAATATA GTATGGCTTT 1620

TTTCATCGTT TGAGATTTTA TTTTTGCTTG TTAATTTTAG AACATTTATT CAGTTACCAG 1680

TGGATATTTT TATTGGTATG ATAATATTTT TAATGCTGTG GATATTTATT ATGTTAGGTA 1740

TAGTGTGTCT TAGTTATTAT ATAACTTTAT TATTTAGCAA GGAGGCTTAG TATGTTTAAA 1800

AAAATAGGTA TAATGAGCAT TTGCATATAT ATAATTATTT TATACTGCTT GAGAATGTAT 1860

CGTATTATCA ATAATATTGA AACAATCTTG CTAACGGTTA TATGCTTAAT GTTATTGTTT 1920

TTTTTAAGAC GTTTATTTGA TAAAGATAAG TAAATAGATG TTAAGTAAAA ATGTAGAATA 1980

TAAAGGAGGT GCAATGAGTA TGATTGAAGT TAGCCATTTA TCAAAAAGTT TTGGTGATAA 2040

AATAGCTTTA AATAATATAA GCTTCACTGT TAAAGAAGGT TAGATTTTTG GATTTTTAGA 2100

ACCATCTGGT TCTGGAAAGA CCACAACGAT TAATATTCTG ACTGGGCAGT TCCTTGCCGA 2160

TAAAGGACAA TCTATTATTT TGGGACAAAA ATCTCAAAAT TTAACAAGCG GTGAATTAAA 2220

GAGAATTGGA TTGGTTAGCG ATACAAGTGG ATTTTATGAG AAAATGTCTC TGTATAACAA 2280

TCTTCTTTTT TATAGTAAAT TTTATAATAT TAGTAAATCA CGTGTTGATA ATTTGTTAAA 2340

GCGAGTAGGA TTATATGATA GTCGCAAGAT GGTAGCAGGA AAATTATCCA CTGGAATGAG 2400

GCAACGAATG CTTTTAGCAC GAGCTCTTAT CAACAACCCC GCTGTACTCT TTCTGGATGA 2460

ACCGACCTCA GGTCTAGATC CCACAACTTC TCGAACAATT CATGAGTTAA TTTTAGAATT 2520

GAAAACAGCA GGGACAACGA TTTTTCTAAC GACTCATGAT ATGAATGAAG CAACTCTTTT 2580

ATGTGATTAT GTTGCCTTAT TAAATAAAGG GAAATTAGTT GAGCAAGGAG CTCCTTCTGA 2640

ACTCATTCAA AGATATAATA AAGATAAAAA GATTAAGGTT ACAGATTATA ATGGGAATCA 2700

GATAACTTTT GATTTTACAT CACTAGAACA GGTATCTCAG ACTGATCTGG AAAATATTTT 2760

TTCAATTCAT TCATGTGAGC CTACTTTAGA AGATATTTTT ATCACATTAA CAGGAGGAAA 2820

GCTAAATGCT TAAACGGTTT CTGGCTTTGG TATGGTTGCG TTGTCAAATC ATCCTTTCCA 2880

ATAAGAGTAT TTTATTGCAA GTTTTAGTGC CTTTTGCTTT CACATATTTT TATAAATATC 2940

TTATGGAAAC ACAGGGGAAG GTCAACGATC AACAGGCATT AGTTCTTTTG ATGATGTGTT 3000

TACCTTTTTC TTTTTCTTTG GCTGTTGGAA GTCCTATAAC TATTATCTTG TCTGAAGAAA 3060

AAGAAAAGTA CAATTTACAA ACTCTTCTGT TGAGTGGTGT TAAAGGCTCC GAATACATTT 3120

TATCAACTAT GTTTCTTCCT TTTTTGCTAA CTTTTGTGAT TATGGGAACT ACTCCTCTTA 3180

TTTTAGGAGT TACAATTGTA CATACTTTTA ATTATATTAC AATCGTTCTT CTAACCTCTT 3240

TATCCATCAT TTTATTCTAT TTATTGATAG GTTTAACCGC GAAGAGCCAA GTAGTAGCTC 3300

AGGTTATCAG TCTTCCTGCT ATGATTTTAG TTGCTTTCTT ACCGATGCTA TCTGGTTTGG 3360

ATAAGACAGT TGCGAAGATA ACAGATTATA GTTTTATGGG ACTATTTACT AAGTTTTTCA 3420

CAAAATGGGA GGAATTTTCA TGGAATAAAA CTCTAATTCC TAATCTAACA CTACTTATTT 3480

GGATTGTTCT TCTATTAACT TTAATTACGA TAACTATTAG GAAAAAGAAA ATTTCTTAAT 3540

TGAGTTATTT TAATGATTAT AAACACAAGT GGGAAGGAAA AAATGAACTG ATCTTTTTGA 3600

CAGCAATTCT ACAGAATAGT CTTATTGCTA TATTTTGATT TGAGTGTACG AAAAAAGAAA 3660

AATAACAATA GTGCTCATAC TAATTGCAGA AGTTTTGGGT GATAAGATAA CTGATAAATT 3720

GCAATAAAAA ATGCAACATT TTTAAATCTC CTCTATAAGT GCTTCAAAAA GTGCTTCAAA 3780

ACCTGTCTTG TAATCCAAGT ATTTTTGGGG ACGGTGATTA ATAAGCTAGC AAAGCATCAT 3840

TAAGGATTTT TTCGGTAATT GTTGCCAAAT CGGTTTAAGA AAATACTCAC GAAGAAGTCC 3900

ATTCGCATTC TCATTACTTC CCCTTTGCCA AGATGAATAG GCATCCGCAA AATAAAACAG 3960

AATTCCCATT TGTTCAATTA AAGGGTAACA AGCAAACTCT TTTTCTCTGT CCGAAGTGAA 4020

AGTCTTTAAC TATTCTTTTG GAAAGAGTCT TGTGAGGTGT TCAATAGCAG TCAACATGGA 4080

TTTAGCTGTT TTTACTTGAC AAGTGCTAGT AGAAATAATA GAATAGTAAA AAACCTTTAA 4140

AGCAGTCCAG AGAGGCAGCT AAGGTTAGAC GGTGAAAGGG TGGAGACTAC CCATTTTTCG 4200

TGGAACCTTG CTGTTGGCAG GTTCCTTTTT TCGTGGCTTC TGTTGGCCAG ACTCTCTCAC 4260

TAGTAAAGGT AAAAGGAGAA ACCTATGCGA GAACATCGTC CAATCATTGC TCTTGATTTT 4320

CCTAGTTTTG AGGCGGTCAA GGAATTTTTA GCTCTTTTCC CAGCAGAAGA AAGCCTTTAT 4380

CTCAAGGTAG GGATGGAGCT TTATTACGCA GCGGGGCCTG AGATTGTGTC CTACTTAAAA 4440

GGTTTGGGTC ATAGTGTCTT TTTGGATCTC AAACTTCATG ACATTCCTAA TACAGTCAAG 4500

TCAGCCATGA AGATCTTGTC TCAGCTTGGT GTCGATATGA CTAATGTCCA TGCGGCTGGT 4560

GGTGTAGAGA TGATGAAGGC GGCGCGTGAA GGTCTTGGGA GTCAAGCCAA ATTGATCGCT 4620

GTAACTCAGC TCACATCAAC GTCAGAAGCT CAGATGCAGG AGTTTCAAAA TATCCAAACC 4680

AGTCTGCAAG AGTCTGTGAT TCACTATGCC AAGAAGACAG CTGAAGCTGG CTTGGATGGT 4740

GTTGTTTGCT CGGCTCAGGA AGTACAAGTC ATCAAGCAGG CTACCAATCC AGATTTTATC 4800

TGTCTGACAC CAGGGATTCG TCCAGCTGGT GTTGCAGTTG GAGATCAAAA ACGAGTCATG 4860

ACACCTGCTG ATGCCTATCA AATCGGCAGT GACTATATCG TAGTGGGACG TCCCATTACC 4920

CAAGCTGAGG ATCCTGTTGC AGCTTATCAT GCCATCAAGG ATGAATGGAC ACAGGACTGG 4980

AATTAAAGAA CTAGATTAGA AAAATAAAAG GAGAATACCA TGACACTTGC TAAAGATATC 5040

GCTAGCCACC TCTTGAAAAT CCAAGCCGTT TACCTCAAAC CAGAGGAACC CTTCACTTGG 5100

GCATCTGGTA TCAAGTCACC GATTTACACT GATAATCGTG TGACACTAGC CTATCCAGAA 5160

ACTCGTACCC TAATTGAAAA TGGTTTTGTG GAAGCTATCA AAGAAGCCTT TCCTGAAGTA 5220

GAAGTGATTG CAGGAACTGC AACAGCAGGG ATTCCACACG GAGCCATTAT TGCTGATAAG 5280

ATGGACTTGC CTTTTGCCTA CATCCGTAGT AAACCAAAAG ACCACGGAGC TGGTAATCAA 5340

ATCGAAGGTC GCGTAGCTCA AGGTCAAAAA ATGGTAGTGG TTGAAGACCT TATTTCAACG 5400

GGTGGTTCAG TTCTTGAAGC TGTAGCAGCA GCCAAGCGAG AAGGAGCAGA TGTACTTGGA 5460

GTTGTAGCGA TTTTCAGCTA CCAATTGCCA AAAGCAGATA AGAACTTTGC AGATGCTGGT 5520

GTTAAACTTG TGACGCTTTC AAACTATAGC GAGCTTATCC ATCTAGCCCA AGAAGAAGGT 5580

TACATCACGC CAGAGGGCCT TGATCTTCTA AAACGCTTTA AAGAAGACCA AGAAAATTGG 5640

CAAGAAGGTT AGGTCAGTAA GATAAAGAGA GACGAGGCTA CCGAGTCTCT TTTACCATTT 5700

TATTTAAAAT ATGACAG 5717

(2) INFORMATION FOR SEQ ID NO: 103:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5558 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

CCTGGACTTT CTAAAATGAA ATCTTGCGAC CTGGATCAAG CCCTTCATGA GCATTTTTCA 60

GAAGAAGAAT TAGCTGGTCA CTTTCATGTC CTTCTATGGA CTTTTTTTAC AATGGCATTG 120

CTATCACACC CAATACCTAT CTAAGCGCCT GGTTCGTAAA CTTTATTGCA GCTCTTCCTC 180

TAAATTTCCT AATTGTTGAA CCAATTGCCC GTTTTATACT AAGTTCTTTT CAGAAACCAT 240

TTACTGGGGA AGAAGTTGAA GATTTTCAAG ATGATGATGA AATCCCAACT ATTATCTAAG 300

CCAGTTCTGT AAACTACTAA TATTTGAAAT CCACTTCCTT TTAGGGTGCA ATGGTTATAA 360

ATGAATTTTT GAGAGGATCA GAATGAAAAA ACTAGCAACC CTTCTTTTAC TGTCTACTGT 420

AGCCCTAGCT GGGTGTAGCA GCGTCCAACG CAGTCTGCGT GGTGATGATT ATGTTGATTC 480

CAGTCTTGCT GCTGAAGAAA GTTCCAAAGT AGCTGCCCAA TCTGCCAAGG AGTTAAACGA 540

TGCTTTAACA AACGAAAACG CCAATTTCCC ACAACTATCT AAGGAAGTTG CTGAAGATGA 600

AGCCGAAGTG ATTTTCCACA CAAGCCAAGG TGATATTCGC ATTAAACTCT TCCCTAAACT 660

CGCTCCTCTA GCGGTTGAAA ATTTCCTCAC TCACGCCAAA GAAGGCTACT ATAACGGTAT 720

TACCTTCCAC CGTGTCATCG ATGGCTTTAT GGTCCAAACT GGAGATCCAA AAGGGGACGG 780

TACAGGTGGT CAGTCCATCT GGCATGACAA GGATAAGACT AAAGACAAAG GAACTGGTTT 840

CAAGAACGAG ATTACTCCTT ATTTGTATAA CATCCGTGGT GCTCTTGCTA TGGCTAATAC 900

TGGTCAACCA AACACCAATG GCAGCCAGTT CTTCATCAAC CAAAACTCTA CAGATACCTC 960

TTCTAAACTC CCTACAAGCA AGTATCCACA GAAAATTATT GAAGCCTACA AAGAAGGTGG 1020

AAACCCTAGT CTAGATGGCA AACACCCAGT CTTTGGTCAA GTGATTGACG GTATGGATGT 1080

TGTGGATAAG ATTGCTAAGG CCGAAAAAGA TGAAAAAGAC AAGCCAACTA CTGCTATCAC 1140

AATCGACAGC ATCGAAGTGG TGAAAGACTA CGATTTTAAA TCTTAAAAAC CAAAAAAATA 1200

CAGTATCCAC ATTCGGTACT GTATTTCTTT TACTCTCATT CTTAAGTTAA ATTATTAAAA 1260

TCCCATATTT GGTCTATCCA GCCTTCATAA AAGTCTGGCT CGTGGCAGAC CATAAGGATA 1320

GATCCCCTAT ATTCTTTGAG AGCGCGTTTG AGCTCATCCT TTGCATCCAC ATCCAAATGG 1380

TTGGTCGGCT CGTCCAGCAC TAAAACGTTG TTTTCACGAT TCATCAAGAG ACAGAAACGA 1440

ACCTTGGCTT GCTCTCCCCC TGATAATACT TGAATCTGGC TTTCAATATG TTTGGTTGTC 1500

AAACCACAAC GGGCAAGGGC TGCACGGACT TCTGCTTGAT TAAGGGCAGG AAAGGCATTC 1560

CAGACAGCTT CAAGAGGAGT TTGGCGATTA CCGCCTTCTA CTTCCTGCTC AAAATAACCA 1620

AGTTCTAAAT AATCTCCACG CTCCACTTCC CCAGCGATTG GCGAGATAAT GCCCAAGAGA 1680

CTCTTCAAGA GAGTTGTTTT TCCAATACCA TTAGCACCAA TAATCGCAAC CTTTTGATTG 1740

CGTTCGAAGG TAAGATTTAA AGGCTTAGTA AGAGGACGGT CGTAACCAAT TTGCAAGTTC 1800

TTGGCTTGGA AGATAAAGCG CCCTGGTGTA CGAGCTGGTT TGAAATCAAA GGATGGTTTT 1860

GGTTTCTCAC TTTGGAGTTC GATAATATCC ATCTTATCCA ATTTCTTTTG ACGAGACATA 1920

GCCATATTAC GAGTTGCAAC ACGGGCTTTA TTACGAGCCA CAAAGTCCTT GAGGTCTGCA 1980

ATCTCTTTCT GCTGGCGTTC GTAGGCTGCC TCTAGCTGAG ATTTCTTCAT AGCATAAACT 2040

TCTTGGAACT GGTAGTAGTC ACCAGAGTAA CGCGTCAGCT GTTGATTTTC CACATGATAG 2100

ACAATATTAA TAACGTCATT GAGGAATGGA ATATCGTGCG AAATGAGAAC AAAGGCATTC 2160

TCATAGTTTT GGAGATAGCG CTTGAGCCAA TCAATATGCT CAGCATCCAA GTAGTTGGTC 2220

GGCTCGTCCA ACAGCAAGAT ATCAGGCTTT TCAAGGAGAA GTTTTGCCAA AAGCACCTTG 2280

GTTCTTTGCC CACCTGACAA AGAAGTTACA TCCGTATCCA TGCCAAAGTC CATAACACCA 2340

AGAGCACGCG CTACTTCGTC AATCTTAGCA TCCAAGGTAT AGAAATCACG ACTCTCCAGA 2400

CGGTCTTGAA GTTCTCCTAC TTCTTCCATG AGAGCATCAA CATCCGCGCC GTCTTCAGCC 2460

ATTTTCATAT AGAGGTCATT GATACGAGCT TCAGCTTTGA AAAGCTCATC AAAAGCCGTA 2520

CGGAGAACAT CACGCACCGA CTGTCTTTCA GCAAGGACAG AGTGCTGATC CAAGTAACCA 2580

GCCGTCACAT ATTTGGACCA CTCAACCTTT CCTTCATCTG GCAGCATTTT ACCAGTCACG 2640

ATACTCATAA AGGTTGATTT TCCTTCACCA TTGGCACCGA CCAGGCCGAT ATGTTCTCCC 2700

TTGAGGAGAC GGAAGGACAC ATCTTCAAAA ATTGCACGGT CACCAAAACC GTGACTCAGA 2760

TTTTTAACTT CTAAAATACT CATTTTAATT CCTTACCTTG TTTTTATGTA ATCGTTTATA 2820

AAGGAGCCAA GCCAGATAGC CACCCAAAGT GTTGGTCCAC AAATCATCAA TCTCAAAGAC 2880

GCGATTGAAA TCAAAGAAAA AGTCCAAGAT TAATTGCGTA CACTCGATTC CAAGACTCAC 2940

AAGAAAACTA AAAAGAAGGA CCTTTTTTGT TTTCCGCAAA TTTGGAAATA GATAAAGGAG 3000

TTGGAAAATC AGAGGAAAAA ACAAGAAGAC ATTGAGGATA TTTTGTAAAA AAATCCAACA 3060

TAATTGTCCA ATGTCACTCA CTTCGCCCAG TTTCCAGAGA GAATTGAAAG GAGTCAAAAG 3120

AAAAACCAGG CGTCCAAGAT GCTGAATACC TGGAGTTCCC ACTCCCACGG TAGATTGTTC 3180

TTGAGGAGTA AAGCAAAAAC AGACAATGCA AATGCTATAG AAAATGACTC CCCAGACCAA 3240

AATATGATTA TAAGTCTTCT TCATCATTAA GGATTTACCG CTGCGACTGC CTTCTGGCGG 3300

TCACGTTTCA TTGTGTTAGA GCGCAATTGT CCACAAGCTG CGTCAATATC TGTACCATGC 3360

TCTTGACGAA CCACACAGTT GACCCCTTTT TTCTTAAGCG TATCATAGAA AGCCAACACG 3420

CACTCTTTGG GACTACGGCT ATATTGGTCA TGCTCACTAA CTGGGTTATA AGGAATCAAG 3480

TTTACATAAG ACAATTTCTT GATGTTCTTG AGCAATTCAG TCAATTCCAA GGCTTGTTCT 3540

ACACCGTCGT TGACTTCATT AAGCATGATA TATTCAAAGG TTACACGACG GTTTGTTGTC 3600

TCAATGTAGT ATTCAATAGC AGCAAAGAGT TTTTCAATCG GAAAGGCACG GTTAATCTTC 3660

ATGATACTTG AACGAAGTTC ATTGTTAGGT GCGTGAAGAG ACACGGCAAG ATTGACCTGA 3720

ACCCCTTCAT CAGCAAAGTC ACGAATTTTA TGAGCCAAAC CTGAGGTTGA AACCGTGATG 3780

TGACGAGCAC CGATAGCCAT TCCTTTATCA TCATTGATAG TACGAAAGAA ATTCAAGACA 3840

TTGTTGTAAT TATCAAAGGG CTCACCGATT CCCATGACAA CGATATGGCT GATGCGTTCA 3900

TCCTGACCAC GCTCATCAAA GTATTTCTGA ACCAGCATGA TTTGCGCTAC GATTTCACCG 3960

TTATTGAGGT CACGTTGCTT CTTAATCAAA CCAGAGGCAC AGAAGGTACA ACCGATATTA 4020

CAGCCGACCT GAGTGGTCAC ACAGACAGAT AAACCATAGT GTTGACGCAT GAGTACAGTC 4080

TCAATTAACA TACCGTCGGG CAATTCAAAG AGATATTTGA CTGTACCATC AGCAGACTCT 4140

TGCACAATAC GTTGTTTCAA GGGATTGACC ACAAACTGGT CATTGAGCTT AGCAATCAAA 4200

TCCTTGGAAA GGTTGGTCAT TTCTTCAAAT GACTGCACAC GTTTACGGTA GAGCCATTCC 4260

CAGATTTGAT CTGCACGGAA TTTCTTTTCT CCCTGCTCCA ATACCCATTC CTGCATGGTT 4320

TGATGTACCA AACTATGAAT TGAGGGTTTC ATTTCTTCTC CTTATTCTCT ACTCACTTCT 4380

GACGAATGAC AAAATGACGT TGTCCCTTGT CGTCTTTCTG ACGACGTCTA TTTTTCTTAT 4440

CTGCATTCGA CTTTCGTTTA GTTTGAGTCG GTTTCTTTCC TTTTCTAGAA GGTGTTTCTT 4500

CTTCCGTCTT ACGCATTTTC TTGTCAAATG ATGCTCGCTT AGGGGCTTCA TTTTCTAAGA 4560

CAAAATAGGC ACAACCATAA CTACAATACT CTAAAAGGTA GTCTTGTAAA CGACTGATTT 4620

TTTCAAGTTT TTCTTCTGTT CGGTCATCCT TGTAAAAACC TCGTAGGCGA AGCTGTTCGT 4680

TGCTCCAGTC CCCCACGATA TAATCAAACT TGGTTAATAC TTCTGAAAAA CGCTGATTAA 4740

AAGTCGTCAC ATCAAAGGCA TCCTTGATAT TTTCAACCAA GGAAAAAGCT ATCCCTTCCG 4800

TTTCGACCTT GTCCCCGTGT AAATGGAACT CCGGACCAGG AAACTTGTTA TAGTTGTATA 4860

ATTCAGGTGC AATTTCTTTT CGCATAGATA TCCTTTTTTC ACGATTACTT AATACTTTAT 4920

TCTACCATAA TTTCTAGCAG TTAGCACGTT TCTCATAAAA ATGAAAAAAG TCTGACGATT 4980

TTGTCAGACC AGAATCTTAT AACCTAAAAA GAGAAGAACA ATTCTTCCCT CCAACTATCA 5040

TTATTTAGCA GCTGCGTACA ATTCATCTAC TTTATTCCAG TTGATTACTG AAAAGAAAGC 5100

TTTGATGTAG TCAGGACGCA CGTTGCGGTA TTTCACGTAG TAAGCATGTT CCCAAACGTC 5160

CAAGCCCAAG ATTGGTTTTT TACCTTCTGA GATTGGTGTG TCTTGGTTTG CTGTTGAAGT 5220

CACTTCAAGT TTCCCTTCTT TGTTGACAAC CAACCATGCC CAACCTGAAC CAAAACGAGT 5280

TGTTGCTGCT GCAGTGAAGG CTGCTTGGAA TTCTTCAAAT GAACCAAATG TTGCATCGAT 5340

TGCTGCTGCC AGTTCTGCTG AAGGAGCTGT TTTCTCGGGA GTCATCAATT CCCAGAAAAG 5400

AGCGTGGTTC AAGTGTCCGC CACCATTGTT GATAAGTGCT TGACGGATAT CAGCTGGGAT 5460

AGATTCTACA TCAGCAAGCA AGGCTTCAAG GTCTTCACCG ATTTCAGGGT GTTTTTCTAA 5520

AGCTGCATTG GCATTGTTGA CATAAGTTTG ATGGTGTT 5558

(2) INFORMATION FOR SEQ ID NO: 104:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6735 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

GGAATTGTAA ATATCATATT GTTTTTGCAC CCAAATATCG TCGTCAAATC ATTTATGGCA 60

GATACAAAGC TAGTATCGGA AGAATCATAC GTGACTTATG TGAGCGTAAG GGTGTAATAA 120

TCCATGAAGC GAATGCTTGT TCAGACCATA TTCACATGCT TATCAGTATT CCTCCGAAAC 180

TTAGTGTTTC GTCCTTTATG GGCTATTTAA AGGGCAAGAG CAGTTTGATG ATTTTTGATA 240

AGCATGCGAA TTTAAAATAC AAATATGGCA ATCGCAAGTT TTGGTGTAGA GGCTATTATG 300

TAGATACGGT AGGCCGTAAT CAGAAAGTGA TAGCTGAATA TATTCAGAAT CAATTACAAG 360

AAGACAGAGT AGCAGACCAG CTCACGTTAT TCGAGTCAGT AGATCCGTTT ACTGGCGAAA 420

TAAATAAGAG GAAGTAACTA AGGTGCTTTA GCACCTGCTC GGGAAAGTGG TGCGCGAGGA 480

AGCTATTTCG GTGGGCCTTT GGCCCTGGCC GGTAGAAGCG GCTTATAGCC GCAGAACAAA 540

CCACCAGTTC ACACTGGTGG TTTTGATTTA AAAAACTTGA TACATAAAAA TAAAAGTCTA 600

TATAAAGGAT GGTAAAATTC CTGTTGTCCG ATTTGGACAA TATCCTAAAT AGTTACAATA 660

TATGGTCTAT ACTTTTTCTT AGGAGAAAGC TAGATGTACA GACGTTTGAG AGATTTGAGG 720

GAGGATCATG ATCTGCCCCA AAAGCAAATA GCTACAATAC TTTCGTTTAC AAATTCAGCT 780

TATGCCAAAA TTGAACGGGG TGAGCATGCG TTGACGGCTG ATGTATTGGT TAAACTCTCA 840

GATTTCTATG ACGTCAGTAC AGACTATTTA TTGGGATTAA CTGATTTTCC TGATAAAATT 900

CGCTTTAGAA AATAATCTCC TCAATTTCAT AGAGTTTGAA AATGAGTGAG ATTTTTTATT 960

TGCCCTTTGA CAACTGAATA GCCTAAAATG GTACTTTCCT CATTTGTGGA GCAAATTTGA 1020

ATGGCTCGCC ATGATAAGAG CGATTTTAAA ATCATCAATA AAATAGAGCG ATACTTTATA 1080

TGCCATGATA CAAATGATAT ACAATGATAC TTCTGACCGT TCAGCCTGCC AACGTAAAAG 1140

AGCAGCAAGT GAAATTCTTA TGATGACTTC ATCAGTCATG CCACGTTGAA TGTGTGAGTT 1200

TGTTAGATAA ACGCAATTAA TCCTCAAAAG GTTCCCCGAA CCTTTTGAGT TCTACAGACG 1260

CATCACGTGG AGTGTGTAAG CTTGTTGCTA AAAGCGTAAA AACCTTGGAA CGAAAGGAAT 1320

AATAGACTTT CTGCGAAACA AAAATATAAT ACAATAAAAC TATGAATGAT GAAGCAAGTA 1380

AACAATTGAG CGATAGCCGT TTCAAGATCC TTGTAGGTGT TCAGCGCACG ACTTTTGAAG 1440

AGATGTTAGC TGTGTTAAAA ACAGCTTATC AACGTAAACG CGCAAAAGGT GGACGAAAAA 1500

GCAAATTAAG CCTAGACGAT CTCCTTATGG TAACTATTCA ATACATGCGA GAATAGAGCA 1560

CTTATGAACA AATTGCGGCT GATTTTGGCA TTCACGAAAG CAACTTAATC CGTCGGAGTC 1620

AATGGGTTGA AGCAACTCTT ATTCAAAATG GTTTTACGAT TTCAAATTCT GCCTTAATTC 1680

TGTAAAAACA GTAAAATTCG AAGGATTGTA AGGTAAGAGT TTTTTTCTTT CTGAAAAAAT 1740

GGTATAATAG CAATCAAAAC TAGAAAATAA AACGGAATTT GGAACAGATT TGTCTGTATC 1800

CTAGTAGAGT GGTGATACTA TGAAGATTAG TAAGAGGCAC TTATTAAATT ATTCCATCTT 1860

GATTCCCTAC TTGCTTTTAT CTATTTTGGG CTTGATTGTG GTCTATTCGA CCACCAGTGC 1920

TATTTTAATT GAAGAAGGCA AGAGCGCCTT GCAGTTGGTT CGAAACCAAG GAATCTTTTG 1980

GATTGTTAGT TTGATACTGA TTGCCTTAAT TTATAAATTG AGACTAGATT TTTTGAGAAA 2040

TGAGCGACTA ATCATTTTAG TTATATTAAT AGAAATGCTT TTATTGTTCT TGGCTCGTTT 2100

TATTGGTATT TCCGTAAACG GGGCATACGG TTGGATTTCG GTTGCAGGAA TAACTATTCA 2160

GCCAGCTGAG TACTTAAAAA TCATTATTAT TTGGTATTTA GCTCACCGAT TCTCCAAACA 2220

GCAAGAAGAA ATAGCTACTT ATGATTTTCA AGTTTTGACT CAAAATCAAT GGCTTCCCCG 2280

TGCTTTTAAT GATTGGCGAT TCGTTCTCCT AGTTCTGATT GGAAGTTTGG GAATTTTCCC 2340

TGATTTAGGA AATGCGACTA TTTTAGTCTT GGTTTCCTTG ATTATGTATA CAGTTAGTGG 2400

AATCGCTTAT CGCTGGTTTT CAACCATTCT GGCGCTCGTA TCTGCCGCTT CTGTCTTTGT 2460

CTTGACCACT ATCAGCCTAA TCGGTGTTGA GACCTTTTCA AAAATTCCAG TATTCGGCTA 2520

TGTAGCCAAG CGCTTTAGTG CCTTTTTTAA TCCTTTTGCC GATCGTGCTG ATGCAGGTCA 2580

CCAGTTAGCT AATTCTTATT TTGCCATGGT CAATGGCGGT TGGTTTGGTC TAGGTCTTGG 2640

AAACTCGATT GAAAAACGAG GTTATTTGCC AGAAGCTCAT ACAGACTTTG TCTTTTCTAT 2700

CGTGATTGAA GAATTTGGCT TTGTTGGTGC CAGTCTTATT TTAGCTCTCT TGTTTTTCAT 2760

GATTTTGCGG ATTATCTTGG TCGGTATCCG AGCGGAGAAT CCTTTCAATG CCATGGTTGC 2820

ACTCGGTGTC GGAGGGATGA TGTTGGTTCA GGTATTTGTC AATATCGGAG GGATTTCGGG 2880

CTTGATTCCA TCTACAGGAG TGACTTTCCC CTTCTTATCC CAGGGTGGAA ATAGTCTTCT 2940

AGTCTTATCA GTGGCAGTAG CCTTTGTCTT AAATATTGAT GCCAGTGAAA AACGCGCTAA 3000

ATTGTACCGA GAATTGGAAA ATCAACCAAT GAACCTTCTG TTGAAGTAGG ATAAAGAAAG 3060

GATAGTTTAT GTCTCTTCAA AAATTAGAAA ATTATAGTAA TAAAAGTGTT GTGCAAGAAG 3120

AAGTCTTGAT TCTAACAGAA TTACTGGAAG ATATTACTAA AAATATGCTT GCCCCAGAGA 3180

CCTTTGAAAA AATAATACAG TTGAAAGAAT TATCAACGCA GGAAGATTAT CAAGGTCTAA 3240

ACCGTCTAGT GACTAGCTTA TCAAATGATG AAATGGTCTA TATTTCACGC TATTTCTCTA 3300

TCTTGCCTCT TTTGATTAAT ATTTCAGAGG ATGTGGATTT AGCTTATGAA ATCAATCATC 3360

AAAATAATAT TGATCAGGAC TATTTAGGTA AATTATCTAC AACGATTAAA TTGGTAGCAG 3420

AAAAGGAAAA TGCCGTTGAG ATCCTAGAAC ACTTGAATGT TGTCCCTGTT TTGACAGCCC 3480

ATCCAACACA AGTGCAACGC AAAAGTATGT TGGATTTAAC AAATCATATT CATAGTCTTT 3540

TGCGTAAATA CCGTGATGTT AAGTTGGGGT TGATCAATAA AGATAAATGG TACAATGATT 3600

TGCGTCGTTA CATCGAAATT ATCATGCAGA CAGACATGAT TCGTGAGAAA AAATTAAAAG 3660

TGACTAACGA AATCACGAAT GCTATGGAAT ATTATAACAG CTCCTTTTTG AAAGCTGTAC 3720

CTCATTTGAC GACGGAGTAT AAGCGCTTAG CGCAAGCGCA TGGTCTGAAT TTAAAACAGG 3780

CTAAACCAAT CACCATGGGT ATGTGGATAG GTGGTGACCG TGATGGAAAT CCATTTGTTA 3840

CAGCAAAGAC CTTGAAGCAG TCTGCACTCA CTCAGTGTGA AGTCATCATG AACTACTATG 3900

ATAAAAAGAT TTACCAACTT TATCGTGAAT TTTCTCTTTC AACTAGCATT GTCAACGTCA 3960

GCAAGCAAGT CAGAGAAATG GCTCGTCAAT CCAAGGATAA CTCGATTTAC CGCGAAAAAG 4020

AGCTTTACCG TCGTGCCTTG TTTGATATTC AATCAAAAAT TCAGGCAACT AAAACCTATC 4080

TGATTGAGGA TGAAGAAGTT GGGACTCGTT ATGAAACCGC CAATGATTTC TACAAGGATT 4140

TGATTGCCAT TCGAGATTCT CTACTAGAAA ATAAGGGCGA GTCCTTGATT TCAGGTGATT 4200

TTGTGGAATT ATTGCAGGCA GTAGAGATAT TTGGTTTTTA CTTAGCATCA ATTGATATGC 4260

GACAAGACTC TAGCGTCTAT GAAGCCTGTG TGGCAGAACT CTTGAAATCA GCAGGAATTC 4320

ATTCTCGTTA TAGCGAGTTG AGCGAAGAAG AAAAGTGTGA CCTTCTCTTG AAAGAATTAG 4380

AAGAAGATCC CCGAATTCTT TCTGCGACTC ACGCAGAAAA ATCAGAATTA TTAGCAAAAG 4440

AATTAGCTAT TTTTAAGACG GCTCGTGTTT TGAAAGATAA GTTGGGAGAT GATGTCATCC 4500

GTCAGACCAT CATTTCACAT GCAACCAGCC TTTCTGATAT GCTAGAATTA GCTATTCTGT 4560

TAAAAGAAGT AGGACTGGTG GATACGGAAA GGGCGCGTGT TCAGATTGTT CCCCTTTTTG 4620

AAACAATTGA AGACTTGGAT CATTCAGAGG AAACAATGAG AAAATATCTT TCTCTTAGCC 4680

TTGCCAAAAA ATGGATTGAC TCACGAAATA ACTACCAAGA AATCATGCTT GGCTACTCTG 4740

ACAGTAATAA AGATGGCGGT TACTTGTCAT CATGTTGGAC CCTCTACAAG GCTCAACAAC 4800

AATTGACTGC TATTGGAGAT GAATTTGGCG TTAAGGTTAC CTTCTTCCAT GGTCGTGGTG 4860

GTACTGTCGG TCGTGGTGGT GGGCCAACCT ATGAAGCCAT TACATCTCAA CCGCTCAAGT 4920

CTATCAAGGA TCGTATCCGC TTGACGGAGC AGGGTGAAGT AATTGGGAAT AAATACGGTA 4980

ACAAAGACGC CGCTTACTAT AACCTTGAAA TGCTAGTATC GGCAGCTATT AACCGTATGA 5040

TTACTCAGAA GAAGAGCGAT ACCAATACCC CAAATCGTTA TGAAACCATT ATGGATCAAG 5100

TAGTGGACCG TAGTTACGAT ATCTACCGTG ATTTGGTCTT TGGTAATGAG CATTTCTATG 5160

ATTATTTCTT CGAGTCAAGT CCAATCAAGG CTATTTCAAG TTTTAATATT GGTTCTCGTC 5220

CAGCCGCTCG TAAGACTATT ACTGAAATCG GTGGTTTGCG TGCCATCCCT TGGGTATTCT 5280

CATGGTCACA GAGTCGTGTT ATGTTCCCTG GATGGTACGG GGTTGGTTCA AGCTTCAAGG 5340

AATTTATCAA TAAAAATCCA GAGAATATTG CTATCTTACG AGATATGTAC CAAAATTGGC 5400

CTTTCTTCCA ATCGCTTCTT TCAAATGTTG ATATGGTTTT GTCAAAATCA AATATGAATA 5460

TTGCTTTTGA ATATGCTAAA CTTTGTGAAG ACGAGCAAGT TAAGGCCATC TATGAGACTA 5520

TTTTAAATGA ATGGCAAGTT ACTAAGAACG TTATCTTGGC TATTGAAGGA CATGACGAAC 5580

TCTTAGCTGA CAATCCATAT CTAAAAGCTA GTCTGGATTA CCGTATGCCT TACTTTAATA 5640

TTCTCAACTA TATTCAGTTG GAGTTGATTA AACGCCAACG TCGTGGAGAA TTGTCCAGTG 5700

ATCAAGAACG ATTGATTCAT ATCACCATCA ACGGAATTGC GACAGGATTG CGTAATTCAG 5760

GTTGATAATT TTCAAGAGTG AATGCTAAAA GTGAATATCA AAAAAATTCT AATAGACTAT 5820

TGACAAGTAG TTTAAAAATG ATATAATTTA ACCATTCAGA AAAGTAATCA TACAAACTTT 5880

TTAGAGAGTC TGTGGTAGCT GAAAACAGAT AAGTGGCAAT GATGAAAATT GGGCTGAATG 5940

CTATTTAGAA TTTGAAATTA TAAAAATTCG GTAAGCACAC CTTACAGTGC ATCTCGTTAT 6000

TGCGAGACTG AGCGATAGGG AAATTCCCTA TAATTGAGGT GGTACCGCGC ATCGACGTCC 6060

TCACACAAGT TTTTTGTGTG AGGATTTTTT TGATGGAGGT TAGTATGGAA AGAAAACGAT 6120

GGCGTCGCTT GTTTAGATAA GTGAAATATG TTAAAGGAAA TAAAAAGGAG AAACAGAATG 6180

AAAAATAAAC GTTTAATTGG AATTATTGCT GCATTAGCAG TCTTAGTAGC AGGAAGCTTG 6240

ATTTATTCTT CAATGAATAA ATCAGAAGCT CAGAATAATA AGGATGAGAA GAAAATAACC 6300

AAGATTGGTG TGCTTCAATT TGTGAGCCAT CCATCCCTTG ATTTGATTTA TAAAGGGATC 6360

CAAGATGGAC TTGCAGAAGA AGGATATAAA GATGATCAAG TTAAAATTGA TTTTATGAAC 6420

TCAGAAGGTG ACCAAAGTAA GGTTGCGACA ATGAGTAAAC AATTGGTTGC AAATGGGAAT 6480

GACCTTGTGG TTGGTATCGC AACACCAGCA GCCCAAGGGT TGGCTAGTGC AACAAAAGAC 6540

CTACCGGTTA TCATGGCCGC TATTACAGAC CCAATTGGTG CTAACTTGGT TAAAGATTTG 6600

AAAAAACCAG GTGGCAACGT TACAGGGGTA TCTGACCACA ATCCAGCTCA ACAACAAGTT 6660

GAACTCATCA AGGCTCTGAC ACCGAATGTG AAAACAATCG GAGCTCTTTA CTCAAGTAGC 6720

GAAGACAATT CAAAA 6735

(2) INFORMATION FOR SEQ ID NO: 105:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6516 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

CTAGAGGATC CCAGCAGGTA AATTGGCTTC AGCTGGCAAA AAAGTTGCCC TCGTTGAACG 60

CAGCAAGGCT ATGTACGGTG GAACTTGTAT CAACATTGGT TGTATCCCAA CTAAAACCTT 120

GCTAGTTGCT GCTGAAAAGG ACTTGTCTTT TGAAGAAGTC ATTGCTACTA AAAACACGAT 180

CACTGGTCGC CTCAACGGTA AAAACTATGC GACTGTTGCT GGTACAGGCG TAGATATCTT 240

TGATGCGGAA GCTCACTTCC TTTCAAATAA AGTCATCGAA ATCCAAGCTG GTGATGAAAA 300

GAAAGAACTG ACTGCTGAAA CAATCGTCAT CAACACTGGT GCTGTTTCAA ACGTCTTGCC 360

AATCCCTGGA CTTGCTACAA GCAAAAACAT CTTTGACTCA ACAGGTATCC AAAGCTTGGA 420

CAAATTACCT GAAAAACTTG GAATCCTTGG TGGCGGAAAT ATCGGTCTTG AATTTGCCGG 480

CCTTTACAAC AAACTTGGAA GCAAGGTCAC AGTCCTAGAT GCCTTGGATA CATTCCTACC 540

TCGTGCAGAA CCTTCCATCG CAGCTCTTGC TAAACAATAC ATGGAAGAAG ATGGCATTGA 600

ATTGCTTCAA AATATCCATA CTACTGAAAT CAAAAACGAT GGTGACCAAG TGCTTGTCGT 660

AACTGAAGAC GAAACTTACC GTTTCGACGC CCTTCTCTAC GCAACTGGAC GCAAACCAAA 720

TGTAGAACCA CTTCAACTTG AAAATACAGA TATTGAACTA ACTGAACGTG GTGCTATTAA 780

AGTAGACAAA CACTGTCAAA CAAACGTTCC TGGTGTCTTT GCAGTTGGAG ATGTCAACGG 840

TGGCCTTCAA TTTACTTACA TTTCACTTGA TGACTTCCGT GTTGTTTACA GCTACCTTGC 900

TGGAGATGGC AGCTATACAC TTGAAGACCG TCTCAATGTG CCAAATACTA TGTTCATCAC 960

ACCTGCACTT TCACAAGTTG GTTTGACTGA AAGCCAAGCA GCTGATTTGA AACTTCCATA 1020

CGCTGTTAAG GAAATCCCCG TTGCAGCAAT GCCTCGTGGT CACGTAAATG GAGACCTTCG 1080

CGGTGCCTTC AAAGCTGTTG TCAATACTGA AACAAAAGAA ATTCTTGGAG CAAGCATCTT 1140

CTCAGAAGGT TCTCAAGAAA TCATCAACAT CATCACTGTT GCTATGGACA ACAAGATTCC 1200

TTACACTTAC TTCACAAAAC AAATCTTCAC TCACCCAACC TTGGCTGAGA ACTTGAATGA 1260

CTTGTTTGCG ATTTAAGTTG AGATTTAATC GTATCGAACA GCCCTCTTTG GGCTGTTTTT 1320

ACTTCTGCGG AATCTCAAAT CTGTCTTTCT CCTCTTTTAT GATATAATAG AAACATGAAC 1380

TTAAAAACTA CTTTGGGCCT TCTTGCTGGG CGTTCTTCCC ACTTCGTTTT AAGCCGTCTT 1440

GGACGTGGAA GTACGCTCCC AGGGAAAGTC GCCCTTCAAT TTGATAAAGA TATTTTACAA 1500

AACCTAGCTA AGAACTACGA GATTGTCGTT GTCACTGGAA CAAATGGAAA AACCCTGACA 1560

ACTGCCCTCA CTGTCGGCAT TTTAAAAGAG GTTTATGGTC AAGTTCTAAC CAACCCAAGC 1620

GGTGCCAACA TGATTACAGG GATTGCAACA ACCTTCCTAA CAGCCAAATC TTCTAAAACT 1680

GGGAAAAATA TTGCCGTCCT CGAAATTGAC GAAGCCAGTC TATCTCGTAT CTGTGACTAT 1740

ATCCAGCCTA GTCTTTTTGT CATTACTAAT ATCTTCCGTG ACCAGATGGA CCGTTTCGGT 1800

GAAATCTATA CTACCTATAA CATGATATTG GATGCCATTC GGAAAGTTCC AACTGCTACT 1860

GTTCTCCTTA ACGGAGACAG TCCACTTTTC TACAAGCCAA CTATTCCAAA CCCTATAGAG 1920

TATTTTGGTT TTGACTTGGA AAAGGGACCA GCCCAACTGG CTCACTACAA TACCGAAGGG 1980

ATTCTCTGTC CTGACTGCCA AGGCATCCTC AAATATGAGC ATAATACCTA TGCAAACTTG 2040

GGTGCCTATA TCTGTGAAGG TTGTGGATGT AAACGTCCTG ATCTCGACTA TCGTTTGACA 2100

AAACTGGTTG AGTTGACCAA CAATCGCTCT CGCTTTGTCA TAGACGGCCA AGAATACGGT 2160

ATCCAAATCG GCGGGCTCTA TAATATCTAT AACGCCCTAG CTGCTGTGGC CATCGCCCGT 2220

TTCCTAGGTG CCGATTCGCA ACTCATCAAA CAGGGATTTG ACAAGAGCCG TGCTGTCTTT 2280

GGACGCCAAG AAACCTTTCA TATCGGTGAC AAGGAATGTA CCCTTGTCTT GATTAAAAAT 2340

CCAGTCGGTG CAACCCAAGC TATCGAAATG ATCAAACTAG CACCTTATCC ATTTAGCCTA 2400

TCTGTCCTCC TTAATGCCAA CTATGCAGAT GGAATTGACA CTAGCTGGAT CTGGGATGCA 2460

GACTTTGAAC AAATCACTGA CATGGACATT CCTGAAATCA ACGCTGGCGG TGTTCGTCAT 2520

TCTGAAATCG CTCGTCGCCT CCGAGTGACT GGCTATCCAG CTGAGAAAAT CACTGAAACG 2580

AGTAATCTGG AGCAAGTTCT CAAGACCATT GAGAATCAAG ACTGCAAGCA TGCCTATATT 2640

CTGGCAACTT ATACTGCCAT GCTGGAATTT CGTGAACTGC TGGCTAGTCG TCAGATTGTT 2700

AGAAAGGAGA TGAACTAATG GTTTATACTT CACTTTCCTC AAAAGATGGC AATTACCCCT 2760

ATCAGCTCAA CATTGCCCAC CTCTACGGAA ATCTCATGAA TACtACGGGG ACAATGGAAA 2820

CATCCTCATG CTCAAGTATG TGGCTGAAAA ACTGGGAGCC CATGTGACCG TTGACATCGT 2880

TTCTCTCCAT GATGACTTTG ATGAAAATCA CTACGACATC GCCTTTTTCG GTGGTGGTCA 2940

AGACTTTGAA CAAAGTATCA TTGCAGACGA CCTACCTGCT AAAAAAGAGA GCATTGACAA 3000

CTACATCCAA AACGACGGTG TAGTTCTGGC TATCTGCGGT GGTTTCCAAC TATTGGGTCA 3060

ATATTATGTT GAAGCTTCAG GAAAACGTAT CGAAGGGCTA GGGGTCATGG GACACTACAC 3120

GCTCAACCAG ACCAATAACC GTTTTATCGG TGACATCAAG ATTCACAATG AAGATTTCGA 3180

TGAAACCTAC TATGGATTTG AAAATCACCA AGGTCGTACC TTCCTCTCTG ATGACCAAAA 3240

ACCGCTGGGA CAGGTTGTCT ATGGAAATGG AAACAACGAA GAAAAGGTCG GTGAAGGGGT 3300

TCATTATAAG AATGTCTTTG GTTCCTACTT CCACGGGCCT ATCCTCTCTC GTAATGCCAA 3360

TCTGGCTTAT CGCCTAGTTA CTACTGCCCT CAAGAAGAAA TATGGTCAGG ACATCCAACT 3420

CCCTGCCTAT GAGGACATTC TCAGCCAAGA AATCGCTGAA GAGTACAGTG ACGTCAAAAG 3480

CAAGGCTGAC TTTTCTTAAA CAAAGGAAAA TGATATCAAA GAACTCCGTT ATCTTGTCGG 3540

AGTTTTTTGT CTTTTCTTTT ACCCTTCTCC CTTGCATTTT CTCTCATTTT TTGCCAAAAT 3600

AGAGGGGTAG AAAGAAGGTA GCATATGTCT AAATTACAAC AAATCCTAAC ATATCTTGAA 3660

TCAGAAAAAC TAGACGTCGC TGTCGTATCT GACCCCGTCA CAATCAATTA CCTCACTGGT 3720

TTTTACAGTG ATCCCCATGA ACGCCAAATG TTCCTCTTTG TCCTAGCAGA TCAGGAACCT 3780

CTCCTCTTTG TCCCAGCTCT TGAAGTAGAA CGTGCAAGTA GCACCGTTTC CTTCCCAGTA 3840

GTGGGCTATG TCGATTCTGA AAATCCATGG CAAAAAATCA AACATGCTCT TCCACAACTT 3900

GACTTCAAAC GTGTCGCTGT TGAGTTTGAC AATCTCATCT TGACCAAATA CCATGGTTTG 3960

AAAACAGTTT TTGAGACTGC TGAGTTTGAC AACCTCACTC CTCGTATCCA ACGCATGCGC 4020

CTCATCAAAT CAGCTGATGA AGTGCAAAAA ATGATGGTTG CAGGTCTTTA TGCTGACAAG 4080

GCTGTTCATG TTGGTTTTGA CAATATTTCT CTTGATAAGA CTGAGACAGA TATCATCGCA 4140

CAAATCGACT TTGCCATGAA ACGTGAAGGT TATGAAATGA GCTTTGATAC CATGGTCTTG 4200

ACTGGTGATA ATGCTGCGAA TCCACACGGC ATTCCAGCAG CTAATAAGGT TGAAAATGAT 4260

GCTCTTCTCC TCTTTGACCT GGGTGTTCTG GTCAATGGCT ATGCGTCAGA TATGACTCGT 4320

ACAGTCGCTG TCGGCAAACC AGACCAATTC AAGAAAGATA TTTACAACTT GACTCTTGAA 4380

GCCCAACAAG CTGCTCTTGA CTTTATCAAG CCAGGTGTGA CTGCTCATGA AGTGGACCGC 4440

GCTGCCCGTG AGGTCATCGA AAAAGCTGGT TATGGTGAGT ACTTCAACCA CCGTCTCGGG 4500

CATGGTATCG GTATGGATGT CCATGAATTC CCATCTATCA TGGAAGGAAA CGACATGGTC 4560

ATCGAAGAAG GCATGTGCTT CTCTGTTGAA CCAGGTATCT ATATCCCTGG TAAAGTCGGT 4620

GTTCGTATTG AAGACTGCGG TGTTGTTACC AAGGATGGCT TCAACCTCTT TACAAGCACC 4680

AGCAAAGATT TGCTTTATTT TGATTAAACT ATATAGCCCC TATGCTTTCC TTTCAAAATA 4740

TCTAGGGGCT ATTTTATTGT CATTTTTCTG CTATTATGCT AAAGAAATTG GCTGCAATAA 4800

TCTAACCCTA AGTGTCTGGA ATGATAACGA GGGTGCTCTC CGCTTTTATC AAAGACAAGG 4860

GATGAAACCC CAAGAAACAA CAATGGAAAT GATAATTGAT TAAGAAGTCA TCTATCAAAA 4920

GATGTTAGAA AAAGTTCAAT TTCACTAGAA AATGAGGAAA ATCTCCCCAC AATAAAACGC 4980

ATAGTATCAG GTATTGTGTA CTGACCCCAA ACAGTTAGAC AATTAATTTA TCCGAAGGAT 5040

TTAGTTCTGT ACTGCACAGG ACTAAGTCCT TTTAGTTTTA CCTTAATTCG TTTGTTGTTG 5100

TAGTAATCAA TATAGTCTAT AATGACTTGT TCCAATTGGT TAAGTGATTT AAATGTTTTC 5160

TCATAGCCAT AAAACATTTC GGATTTTAAA ATGCCAAAGA AAGATTCCAT CATACCGTTG 5220

TCTTGGCTGT TTCCCTTGCG TGACATAGAT GCTTGAATTC CCTTATTCTC TAGGAACCGA 5280

TGATAAGAAT CGTGTTGGTA TTGCCAGCCT TGGTCACTAT GGAGAATCGT ATTCTCGTAG 5340

TGCTTCTCTT TGAATGCCTG TTCCAACATT GTTTGTACTT ATTCTAAATT AGGCGAACAA 5400

GAAAGATTAA AAGCAATAAT TTCGCTGTTA AAGCCATCTA AAACTGGTGA TAAGTAAAGC 5460

TTTTGAGTAC TTGCTGGAAT GGCAAATTCA GTCACATCTG TGTAGCACTT TTCCATTGTT 5520

TTAGAGCCTT CAAATTGGGC TTGAATGAGA TTCTCTGCCT TCTTACCAAC GTCTCCTTTA 5580

TGAGAAGAAT ATTTTCGTTT CTTTCGCATT TTAGCTTGTA AATTGAGTAC TTTCATCAAG 5640

CCTTGAACTC TTTTATGATT TACCAGATAA CCACGATTTC TTAGTTCTAA ATGAACCCGG 5700

CGATAAGCAT AATTTCCCTT GTGTTCGATA AAGATGGATT GAATTTCAGT TTTAAGCTCT 5760

TGGTCTTTAT CTGTTTTGTC TAGCTGTTTC AAGTGATAGT AGTAGGTCCA ACGAGCTAGT 5820

TTAATGGCTT CTAGAAGAAG ATCTAACGAA AACTCAGTCA TTAATTCTTG AACAATTTCT 5880

GTCTTTCTTC TTTCTCTTTT TCCTCCTTCA ATCGGAGTTC TCTTAACTTT TTTAGGATGG 5940

CATTCTCCGC TCTCAGGTAC TCTCCCTCTT GTTTTCTCAA CAATAGTATA CCCGTTTTTC 6000

CTGTATTGTG CTAGCCAGTT AAGAAGTATC GTACGACTTG GGAGACCGTA TTCAAGAGAA 6060

ACTCTATCTT TAGTCCAGCC TTCATGTCAG ACTTTATTAA CCCCAATTAT TCACCCCAAA 6120

TCTAAAAACC ATCCAGAATC CTTGCCTTAG CTTAGATCCT GGATGGTTTC TTTTTTCACC 6180

CAATGGGTGT TTTTTACTAG ACAAAAAAGA GTTTCCCCTT TATGGTATAA GTGTAGAAAA 6240

AAACACAAAA AGAAAGGAAA CTCACATGAA CAGTTTACCA AATCATCACT TCCAAAACAA 6300

GTCTTTTTAC CAACTATCTT TCGATGGAGG TCATTTAACC CAGTATGGTG GTCTTATCTT 6360

TTTTCAGGAA CTTTTTTCCC AGTTGAAACT AAAAGAGCGG ATTTCTAAGT ATTTAGTAAC 6420

GAATGAmCAA CGCCGCTACT GTCGTTATTC GGATTCAGAT ATCCwTGTCC AGTTCCTCTT 6480

TCAACTGTTA ACAGGTTATG GAACGGAATA TGCTTG 6516

(2) INFORMATION FOR SEQ ID NO: 106:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14654 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

TTTTCAACCC ATATCGTGGC TCCTGAATAC TACTTACTGA CAACTATGCT ATCAGAGACT 60

TCTCTACTTG TTTTCTATAT CATTTTCATC CATAGAAAAC AACTCATCCA CTTGGGACAT 120

ATCTTTAGCT ATACTGTTCG ATACTCTCTC TTTTCACTTT CCTTTGTAGC AATTTATTTC 180

CTGATTAATT TCGTGTATCC TGTAGATATG GTCATTAATT TGCCATTTTT GATTAATACT 240

GGTTTGATTG TCTTGCTATC AGCTATCTCT TATATTAGTC TACTTGTCTT CACAAAAGAT 300

AGCATTTTCT ATGAATTTTT AAACCATGTC CTAGCCTTAA AAAATAAATT TAAAAAATCA 360

TAGGAGTTTA AAATGAAACA ACTAACCGTT GAAGATGCCA AACAAATTGA ATTAGAAATT 420

TTGGATTATA TTGATACTCT CTGTAAAAAG CACAATATCA ACTATATTAT TAACTACGGT 480

ACTCTGATTG GGGCGGTTCG ACATGAGGGC TTTATCCCTT GGGACGACGA TATTGATCTG 540

TCCATGCCTA GAGAAGACTA CCAACGATTT ATTAACATTT TTCAAAAGGA AAAAAGCAAG 600

TATAAGCTCC TATCCTTAGA AACTGATAAG AACTACTTTA ACAACTTTAT CAAGATAACC 660

GACAGTACGA CTAAAATTAT TGATACTCGA AATACAAAAA CCTATGAGTC TGGTATCTTT 720

ATCGATATTT TCCCTATAGA TCGCTTTGAT GATCCTAAGG TCATTGATAC TTGTTATAAA 780

CTGGAAAGCT TCAAACTGCT GTCTTTCAGT AAACATAAAA ATATTGTCTA TAAGGATAGC 840

CTTTTAAAAG ATTGGATACG AACAGCCTTC TGGTTACTCC TTCGACCGGT TTCTCCTCGT 900

TATTTTGCAA ATAAAATCGA GAAAGAAATT CAAAAATATA GTCGTGAAAA TGGGCAATAT 960

ATGGCTTTTA TCCCTTCAAA ATTTAAGGAA AAGGAAGTCT TCCCAAGTGG TACCTTTGAT 1020

AAAACAATCG ATTTACCCTT TGAGAATTTA AGCCTTCCTG CACCTGAAAA ATTTGATACT 1080

ATTTTGACAC AATTTTATGG AGATTATATG ACCCTACCAC CAGAAGAAAA ACGCTTCTAC 1140

AGTCATGAAT TTCACGCTTA TAAATTGGAG GATTAGGATG CAATATTTAG AAAAAAAAGA 1200

AATTAAAGAA ATTCAACTAG CCCTGCTGGA CTATATTGAT GAGACTTGTA AGAAACATGA 1260

TATTCCTTAT TTTCTCAGTT ATGGAACCAT GCTTGGAGCC ATCCGCCACA AAGGTATGAT 1320

TCCTTGGGAT GATGATATTG ATATTTCCCT TTATCGTGAG GATTATGAGC GTTTACTGAA 1380

GATTATTGAA GAAGAAAATC ACCCTCGCTA CAAGGTTCTT TCCTACGATA CATCTTCTTG 1440

GTACTTCCAT AATTTCGCAT CGATTTTGGA CACTTCTACT GTTATAGAAG ACCATGTTAA 1500

GTACAAGCGT CATGATACCA GCCTTTTCAT CGATGTCTTC CCAATTGATC GATTTACAGA 1560

CTTGAGCATT GTCGACAAGA GCTATAAGTA TGTGGCTCTT CGTCAACTAG CTTATATCAA 1620

AAAATCACGA GCAGTTCACG GTGATAGCAA ACTAAAAGAT TTTCTTAGAT TATGTAGCTG 1680

GTACGCTCTC CGATTTGTCA ATCCTCGCTA CTTTTACAAG AAAATTGATC AACTAGTCAA 1740

AAATGCTGTA ACCAACACTC CTCAATATGA AGGAGGAGTT GGGATCGGTA AGGAAGGGAT 1800

GAAAGAAATC TTCCCAGTTG ATACCTTTAA AGAACTGATT TTAACTGAGT TTGAGGGCCG 1860

TATGTTGCCT GTTCCCAAAA AATATGACCA ATTTTTAACC CAGATGTATG GCGATTATAT 1920

GACACCACCA TCAAAAGAAA TGCAAGAGTG GTATAGTCAT AGCATTAAAG CTTATCGCAA 1980

AAACTGATTG AGGGGGATTA TACAAACTAC TAAGATAGAG GTTATTCAAA AACATAATTT 2040

TAGTAGAAAA TGAAATACAT ATTCCCACAA TAAAACGCAT CATATCAAGG TTTTTGAAAA 2100

ACCTTGATAT GATGCGTTTT ATAATTTTAA AGACTTTTTT CTATAGTAGA TTGAAATAAG 2160

ATGCGAACAA ATCAATTAGA AAATTCAAAT TAATTTATAG AAATATTTTA GTATTCCTGT 2220

GTACTGTTCT AAATTCAGTC TGCTATATCT TATTTTTCTA TTTAAATCGC TTCTGTAACA 2280

AAGCTACGAC TTTCAAGTAC CTTAAGCATG GCATTAGCTG TATCTAGCGC TGTGAAGAGG 2340

GGCACCCCGT GTTCAATGGC TGAACGACGA ATTTGCTCAC CATCTTCGTC AGCAGTTCGT 2400

TTTGTTCCTA CTGTGTTAAT GATAGCTTGA ATTCTTCCTT TGCGTACAAA ACTTGGGATA 2460

TCCTTATCGT CATCACCAAT CTTACCAACA GGTTGGGCTT GCAAGCCATG ACTAGCAAAG 2520

AAGGCTGCTG TCCCTTCTGT CGCAAGGATT CCATAACCAA TGTTTTGGAA ACGACGAGCC 2580

AAGTTCAAGG CTTCTTCTTT GGCATCATCA GCGATGGTAA AGACGACATT ACCAAAAGTT 2640

GGCAAGTGTA GATAAGAAGC TTCAAAGGCT TTATAGAGAG CTTTTTCCAA AGTAGCATCA 2700

GAACCCATAA CTTCACCTGT TGACTTCATT TCAGGACCGA GCAAGCTGTC TACCTTAGCT 2760

AGTTTGGTAA AGGAGAAGAC AGGTGCCTTG ATATGAACAC GGGTGCTTTC AGGGTAAAGT 2820

CCATTTTGGT AGCCAAGTTC TGATAAACTT TGACCAAGAA TGAGTTTGGT CGCTACTTGA 2880

GCCATAGGAA TATTGGTTAC CTTAGATAGG AATGGAACAG TACGGCTGGC ACGTGGATTG 2940

ACCTCAATAA CGTAGACTTT TTCATCCTTG ATAACAAACT GGATGTTCAT CATTCCAAGG 3000

CAGTGAAGAC CGATTGCTAA GCGTTTGGTG TAGTCTGCGA TGGTCTCCTG AACCTTTTGC 3060

GACAAGGTTT GTGGTGGGTA AACAGCCATT GAGTCACCTG AGTGGACACC AGCACGTTCG 3120

ATATGCTCCA TGATACCAGG AATGAGTACA TTTTTACCAT CTGAAATGGC ATCAACTTCG 3180

CACTCTTGCC CAACGATATA AGAGTCGACA AGAACTGGGT GGTCTGGACT AGCCTTAACA 3240

GCAGTTCGCA TGTAAGAACG AAGGTCTTCT TCGTTTTCAA CGATTTCCAT GGCACGTCCA 3300

CCAAGTACAT AAGATGGGCG GACAAGAACT GGGAAGCCAA TCTTGCGAGC TGCAAGAGCT 3360

GCTTCTTCTT CATTGGTAGC CGTTTGTCCT GGTGGCTGTG GAATATCCAA TTCTTTGAGA 3420

GCTTGCTCGA AGAGGTCACG GTCTTCGGCA CGATCTAGGT CAGCAACCTG TGTACCAAGG 3480

ATGGTCACAC CTGCTTTTGC CAATGGCTCC GCAAGGTTGA TGGCTGTTTG ACCACCGAAC 3540

TGAACGATAA CTCCCTTTGG TTGTTCCAAG TCAATGACGT TCATAACATC TTCGAATGTC 3600

AATGGCTCAA AGTAAAGCTT ATCTGATACA GAGAAGTCTG TTGAAACGGT CTCTGGGTTT 3660

GAGTTCATGA TGATAGCTTC ATAACCAGCT GCCTGGATAG CCTTAACAGA GTGAACGGTT 3720

GCGTAGTCAA ACTCAACCCC TTGACCGATA CGGATTGGAC CTGAACCTAG GACAAGTACA 3780

GATTCTTTAT CAGATCTGAT AGATTCATTT TCCCAACCAT AGGTTGAATA GAAATATGGC 3840

GTTTCGGAGT CGAACTCTGC CGCACAAGTG TCTACCATCT TATAAACTGG AACAATCTTG 3900

TTTTCCAAGC GAAGTTGGCG AACTTTATCA TCAGTCGTTC CCCAGAGTTC AGCAATCTTA 3960

CGGTCTGAAA AACCATTAAG TTTGGCTGTT TTCAAAACTT CTAAATCTTG TGGATGAGCA 4020

CCCAATTCTT GCTCAATTTC AAAGATATGC AAGAGTTTAT CAAGATAGAA GATATCAATT 4080

TTTGTAAGCT CTGCAATTTC TTCAGGTGTG TAGCCACGAC GAATGGCTTC TGATACGTAG 4140

AAGAGACGGT CATCTTGGGC TTTGACAACC TTTTCAATCA AGGCATCATC AGAAACTGCT 4200

GCAAGTTCAG GTATTTCATT GTGGTGCACC CCAATTTCAA GGGAGCGGCA GGCCTTGAGA 4260

AGAGATTCCT CGATGTTACG ACCGATTGCC ATGACTTCTC CAGTCGCCTT CATTTGTGTA 4320

CCGAGACGGC GTTCACCCTT TTCAAACTTG TCAAATGGGA AACGTGGAAT CTTAGCAACT 4380

ACGTAGTCAA GGGCTGGTTC AAACATGGCA TAGGTTGAAC CTGTAACTGG GTTTATAACC 4440

TCATCCAAGG TCAAACCTAC TGCAATCTTG GCAGCCAACT TAGCAATCGG ATATCCTGTC 4500

GCTTTAGAAG CAAGGGCTGA CGAACGTGAT ACACGAGGGT TTACTTCGAT AACATAATAC 4560

TTGAAGCTGT TAGGATCAAG AGCTAGCTGA ACATTACATC CACCTTCAAT CTTGAGGGCA 4620

CGAATAATGC TCAAGCTCGC ATCACGAAGC ATTTGGTTTT CATAGTCTGA CATGGTTTGC 4680

GCAGGGGCAA ATACAATGGA ATCCCCTGTG TGAATCCCAA CTGGGTCAAA GTTTTCCATG 4740

TTACAAACAA CCAAGGCATT GTCAGCTGAG TCACGCATCA CTTCGTATTC AATTTCCTTG 4800

AAACCGGCAA TCGAACGCTC AATCAAACAT TGGGTAACAG GTGACAATTT CAAACCATTT 4860

TCAGTGATTT CACGCAATTC TTTCTCGTTG GCACACATAC CACCACCAGT ACCACCAAGG 4920

GTAAAGGCTG GACGAACGAT GACTGGGTAG CCAATTGTCG CTGCAAAGGC AACTGCTTCT 4980

TCTACTGTGT TAACAATTTC AGATTCTGGA ATGGGTTGTT CAAGCTCTTC CATCAATTGT 5040

TTAAAGAGGT CACGGTCCTC CGCTTGGTCA ATGGCAGATA ATTTGGTACC CAGAAGTTCA 5100

ACGCCAAGCT CGTCTAGGAT ACCATTTTTA GATAATTCCA TGGCCATGTT GAGACCTGTC 5160

TGACCACCGA GTGTTGGTAG CAAGGCATCT GGACCTTCCT TACGAAGAAT ACGTGTCACA 5220

AACTCAAGTG TAATCGGTTC AATGTAAACC TTGTCAGCAA TTTCCTTGTC CGTCATGATG 5280

GTTGCAGGAT TTGAGTTAAC CAAAACAACC TCATAACCTT CCTCTTTCAA CGACAAGCAA 5340

GCCTGAGTCC CAGCGTAGTC AAACTCAGCA GCCTGACCAA TAATAATCGG ACCAGAACCA 5400

ATCACCATAA TTTTTTGAAT ATCAGTACGT TTAGGCATAT ATAAGATATT AAGGGTGTCA 5460

AGCGGACAAA GCTAAAATAG GAGTTATGAC GAAGAACTGT CAGTTCTAGG AATAACTATC 5520

TTTTTAGCAC CGTCCGTAGC CCGTATTCAG TTCAGCAAAT ACGGAGCACC CTTCTCCTTT 5580

CTATTCGTCG CCTCTCAGGG CGACATTAAA TAAGATACAA AGGACGAATA GAAAGCGATT 5640

GAATTTTAGG AAATCAAGGA AGGATTGACA ATCCAAGTTG GTTTCTCTAC ATTCTGAGCT 5700

TTCCGTCCGT GTTCAGTTAC ATAAATTCTC CGACGAGCTT TTACTCGTTC TTAGTTTGAT 5760

TGTTTAAAAA CTTCCATCAT CTCGATAAAC TCGTCAAATA GGTAGCTAGC GTCGTGTGGC 5820

CCAGGAGCTG CATCTGGGTG GTATTGAACA GAGAAAGCAG GTTGGTATCT GTGGCGCACA 5880

CCTTCCACTG ACTTGTCATT GATTTCTTCG TGGGTAATAA TCAAGTGCTC TGGCAAATCC 5940

TCGCGGCTGA CTGCATAACC ATGGTTCTGG CTGGTGAAGT CTACTCGTCC TGTTGCGATT 6000

TCACGTACCG CATGGTTGAA TCCACGGTGG CCAAACTTCA TCTTATAGGT CTTAGCCCCG 6060

TTTGCCATTG CAAAGAGTTG GTGTCCCATA CAAATACCAA AGATTGGAAT TTTTCCTTGT 6120

ACACCGCGAA TCATGTCGAG TGCTTGTGGA ACGTCTTCTG GGTTACCTGG ACCATTTGAC 6180

AACATAACTC CGTCAGGATT GAGATGGAGA ATTTCTTCAG CCGTTGTCGA ATAAGGAACA 6240

ACTGTCACGT TACAGTTGCG TTTAGAAAGT TCACGTAGGA TTGAGTGCTT GAGACCAAAG 6300

TCCACTAGCA CCACGCTCAA ACCAACTCCT GGAGCTGGAT AAGACGTTTT AGTAGAAACC 6360

TGTTTGATAT TGTCTGTCGG TAAAACTGTT GCTTGGAGCT GGTCCGTCAC ATGGTCCATA 6420

CTGTCCCCAA CATGGGTCAA GGTTGCACGC ATAGTACCAT GCTTACGGAT AATCTTGGTA 6480

AGAGCACGCG TATCAATTCC TGAAATCCCT GGAATTTTCT TGGCTTTCAA AAATTCATCC 6540

AAGGTCATTT GGTTGCGCCA GTTGCTAGCT CTACGCGCTT CTTCAAAAAC AACGACTCCC 6600

TTACAAGTTG GAATAATGGA TTCATAATCA TCACGATTAA TACCATAATT TCCTACCAAA 6660

GGATAAGTAA AGGTCAAGAT TTGTCCATTA TAAGACTGGT CTGTAATGGA TTCTTGGTAG 6720

CCGGTCATCC CTGTATTAAA GACGATTTCG CCTGTTACAT CAATATCTGC TCCGAAGGCC 6780

TTGCCTTCAA AAACTGTGCC ATCTTCTAAT ACTAGAATTC TTTTTGTCAT ATTTTCACCT 6840

CTCGTGGACG CTCACTGGCG TCTTTTAACG TCTTGTGTTT TAGTTGGCGT TTCTACTCGC 6900

TAGTACGGAT TCTAAGATTG CCATTCGAAC AAAGACACCA TTGGTCATTT GTTGGACAAT 6960

CCGTGATTTT GGTGCTTCAA CCAAGTGGTC TGCTATTTCT ACATCACGAT TGATTGGAGC 7020

TGGGTGCATG AGGATTGCTG TTTCTTTCAA ACGATCGTAA CGTTCTTGAG TCAAGCCATG 7080

TTGGGCATGG TAGTCTTCTT TTGAAAATAC AGCTCCACTA TCATGGCGTT CGTGTTGCAC 7140

ACGGAGAAAC ATCATGACAT CAACCTGATC AATGATTTCA TCAATGGTTA CAAACTGTCC 7200

ATAGTCTGCA AACTCTTGAC TTCTCCATTC CTCAGGTCCA GCGAAAAAGA GTTCAGCTCC 7260

CAAGCGTTTC AAAATCTGCA TATTGGATTT GGCAACGCGT GAGTGGTCCA AGTCACCTGC 7320

AATAGCAACT TTAAGACCCT CAAAGTGGCC AAATTCCTCA TAAATGGTCA TCAAATCAAG 7380

CAAGCTCTGG CTAGGGTGTT GGCCCGAACC ATCTCCACCA TTGATGATGG AAGTCGTAAT 7440

CGTTGGACTA GCAATCAATT CTCTATAGTA GTCGACCTCT GGATGGCGAA TCACACAGAC 7500

ATCCACTCCT AAAGCAGACA GAGTCAAAAT GGTGTCATAA AGTGTCTCAC CCTTATTAAC 7560

CGAGCTAGTC TTCACATCAA AGTCAAGTCG TTCCAATCCA AGTTTAATCT CTGCGACTTC 7620

AAAGGACTTA TGTGTCCGTG TAGAATCCTC AAAGAAGAGA TTGGAAACAA TCGGATGGTC 7680

TTCATAGGGA AGCTGGGCTC CATTTTTAAA CTCAATTCCT CGCTTGATCA ATTTCATTAC 7740

TTGATCGACA GTGAGGTCTT CCATGGACAC CACATGGTTC AATGCTTGTT GATTTTCTGA 7800

CATGGCTACT CCTTTAACTT TCTAAGCTTC TTCAGTAATC AGAACTCTGT CTTGGTCATC 7860

AAGTTCTGTC ATCTCTACGA TGATTTCTTC AGAACGACTG GTTGGGATAT TTTTTCCAAC 7920

GTAATCTGGA CGGATTGGCA ATTCTCTATG TCCACGATCG ACTAGAACTG CTAAACTCAC 7980

ACGCGCAGGA CGACCATGAC CGACAATATT ATCAATAGCA GCACGGATGG TACGACCTGT 8040

ATAGAGCACA TCATCCACCA AGATAACTTC GCGGTCTGTC ACATCGACAG AAACCAAAGA 8100

AGTATCTTCT CCACTTTTAA CATCATCACG GAAAGGTTTA GTATCCAATT CCACAACAGG 8160

AACTGAAAGA TTTTCTAACT GCTTCAAACG TTCTTGGATT CGGTGGGCAA TAAAGACACC 8220

ACGAGTTTTA ATACCAGCCA AGACGATCTT ATTCAAATCT TTGTTGCGTT CGATAATCTC 8280

ATAAGTAATA CGCGTAATCG CTCGTTTGAC GGTCAATTCG TCTACAACTT CTTTTGTTTT 8340

CATGACAAAC CTCCAAAAAG AAAAGTCTCC TTAAACAAGG AGACTTGAAA TTTATAGCCA 8400

AGCGAGCCCT ACTGCACACA GTATAGACTT CACCCTTCTA CTTTATCGCG CTCCTTGCCT 8460

GCCTCACGGG ACAGGTTTAA AGGAATATTT AGTTATCATT TACTATAGCA CAAAGCATGC 8520

TTAAAATCAA GCAAAAAGTT TCAATGTAGC ATCTTACAAA TTGCTAAAAT CATATAATTG 8580

TGGGTACTGG TCACACTCTG GATTTTTTGG ATGGCAAATG GCTCTTCCAA AATAAATCAT 8640

GGCCTGATGG GCAGCTAACC ACTGCTCAGG CGGCAAGATA TCCATGACCC GCTTTTCCAC 8700

CTCAAGTGGC GTCGCTGATT TTTTGACAAT ATCGTGGTGT TTGCAAATAC GCTCCACATG 8760

AGTATCCACT GCAAAGGCTG GAATTCCAAA TCCTACACTC ATGACAACAT TGGCTGTCTT 8820

GCGACCAACA CCTGCCAAAC TCTCCAATTC TTCACGTGTC TGAGGGACTT GACCATCAAA 8880

ATCGTCTAGT AACTGTTGGG CACATTTTTT AAGGAATTTA GCTTTATTCC GATACAATCC 8940

CAAGCGAGAA ATATGTGAAG CAATCTCACT CTCTGTCGCT ACAGACATAG CTTGGGGTGT 9000

TGGAAAGGCA ACAAAGAGAC CTGGTGTGGC CTTATTTACC GCTGCATCTG TCGTCTGGGC 9060

TGATAACATG ACCGCAACCA GGAGTTCAAA ATGATTGGTA AAATCAAGAC TAGGCTTGGC 9120

ATCTGGGAAG AGGGCAATGA TTTCTTCTAG CACCTTTCGT GCTCGTTTTT TTGACAAGAC 9180

CATTATTCAT CTCCGTCAAA TAGTCCTTGT AAGCCAGCAA AAGGACTGTT TTCTTCTTTC 9240

TTTACTGCTT TTTGAGCTTG GTATTCTTCC TCTGTCATGA TTTGCCAGTC ATTTCCTGAG 9300

ATAAATCCTT GACCAGCTTC TTCTTCAGCC GTCAAGACCT TGATAGGAAT GTTTAGCAGG 9360

ATATTGTCTG ATACACTCTC AGCAAGGTCA AGCTCCCCAT TTTCGATGGG CAAGACCAAG 9420

TCATCATCTA AAACTTCTTG ATCTAGCTGG TTAGTTGCGC CTTCCATGAA AACTTCCGTG 9480

ACTGGATAAG ATTCAACTAA CTCAACTGGC TCCATACTGC GACTCGACGC AAGAACAATG 9540

GTATAAGATA GTTGATAATC TAAGAAATAC ATACGGTCTT CATATTGTAC TTTCCCAACT 9600

GCAAGGATAT CTTTTACATC TAAAATTTCT TGATTACGTG CACGCAGGTC ATCAACTAAA 9660

TCTAACGTTT GTTCAAAGTT CAAACCTTCA GACTGCTTAC GAATTTCTTG AATATTTAAT 9720

TTCATACTTC CTCCATAAAG ATTTACTCTC TTGATTATAC CATGAAAAGG CTACAAATCA 9780

GCACACCAAA CTTTGTAATT AAAATTCAAA ATTTTAACAT ATTTACTATG ATAGTTTTAT 9840

TTTTTAGTGC TATACTATAG GGAAAGAGTA CATCAGATCA AGGAGGATGC TCACATGGAA 9900

GACAAGAAAC TCATTCAACT CCTATCCAAG TTAAATAAAA GCTACCAAAA CTGTAAACAG 9960

GGTACGGCAG ATGATATTCG ACTACAAGAG CTGCTAAACA CTACTATGCA AGAGCTCAAA 10020

AAAACGGAAC AGTTGAACAA CAGTATCTTA ATTGATCTTG AGAAATTTTA CCAACCTACC 10080

AGTCTTCTGA TTGGACTGGG TAGCCTAAAA CTAAACGATC AAGCACGCAC TGCTTGGCGA 10140

AACTATGATA AATTCCATTA CGATCATGTC AAACACGTAC TAAGTCTCTA TGGACCTGTT 10200

TTTGAATTTT AGAGCATAGA ATTTCCAGTT TTCTGTTGAC AAAATTTCCT TAAAGGTATA 10260

ATATAAAGAT ACTAATACTC GGAGGTAAGG GAGACATGAA CAACTAAGTC TATCAAATAA 10320

AGAACCTTTA TTTAGTAGAT CTTGTTTTTG TCTCTTTTTG TGTGCTCTTT TATGCTCTTT 10380

TTCTGGCATG TTAATAGAGT TTTTTTGACA TAGACTTTGG GCTCTACTAG GTAAAGTAGA 10440

GCTTTTTGTT ATGCACTATG AACATTCTAG AAAGGGAAAT CATATGATAA AAATCAATCA 10500

TCTAACCATC ACACAAAACA AAGATTTACG AGATCTTGTA TCTGACCTAA CCATGACCAT 10560

CCAAGACGGG GAAAAGGTTG CTATTATTGG TGAAGAAGGA AATGGCAAAT CAACCTTACT 10620

TAAAATTTTA ATGGGGGAAG CTTTGTCTGA TTTCACTATC AAGGGAAACA TCCAATCTGA 10680

CTATCAGTCA CTGGCCTACA TTCCTCAAAA AGTCCCTGAG GACCTAAAAA AGAAAACTTT 10740

ACACGACTAC TTCTTTTTAG ATTCTATTGA TTTAGACTAC AGTATCCTCT ATCGTTTGGC 10800

GGAGGAATTG CATTTTGATA GCAATCGTTT CGCAAGTGAC CAAGAGATTG GCAATCTATC 10860

AGGGGGCGAA GCTTTGAAAA TTCAGCTTAT CCATGAGTTA GCCAAACCCT TTGAGATTCT 10920

ATTTTTAGAT GAACCTTCAA ATGACCTAGA CCTTGAGACA GTTGATTGGC TAAAAGGCCA 10980

GATTCAAAAG ACCAGGCAAA CCGTTATTTT CATTTCCCAT GATGAAGACT TTCTTTCTGA 11040

AACGGCAGAC ACTATTGTTC ACTTGCGACT GGTCAAACAC CGTAAAGAAG CGGAAACGCT 11100

AGTAGAGCAT TTAGACTATG ATAGCTATAG TGAGCAGAGA AAGGCTAATT TTGCCAAACA 11160

AAGTCAGCAA GCTGCTAACA ACCAAAGAGC CTACGATAAA ACCATGGAAA AACATCGGAG 11220

AGTTAAGCAA AATGTAGAAA CTGCGCTTCG AGCTACCAAA GATAGTACTG CCGGTCGCCT 11280

ATTGGCTAAA AAGATGAAAA CTGTCCTCTC ACAAGAAAAA CGCTACGAAA AGGCAGCTCA 11340

GTCCATGACT CAAAAGCCAC TTGAAGAGGA ACAAATCCAA CTTTTCTTTT CAGACATCCA 11400

ACCATTACCA GCTTCTAAAG TCTTAGTCCA ACTGGAAAAA GAAAATTTGT CCATTGACGA 11460

CCGAGTTTTG GTTCAAAAAC TACAACTAAC TGTCCGTGGC CAAGAAAAAA TCGGTATTAT 11520

CGGGCCAAAT GGTGTTGGGA AATCAACTCT GTTAGCCAAG TTACAGAGAC TTCTGAATGA 11580

TAAAAGAGAG ATTTCACTTG GTTTTATGCC ACAAGATTAC CACAAAAAAC TGCAATTGGA 11640

TTTATCCCCA ATAGCCTATC TCAGTAAAAC TGGGGAAAAA GAGGAACTAC AGAAAATCCA 11700

ATCTCACCTA GCTAGTCTCA ATTTCAGTTA TCCAGAAATG CAGCATCAAA TTCGCTCCTT 11760

ATCTGGCGGA CAACAGGGAA AACTCCTGCT TTTGGATTTA GTCCTGCGCA AACCAAACTT 11820

TCTCCTGCTG GATGAACCCA CACGAAACTT TTCTCCCACT TCTCAACCCC AAATCAGAAA 11880

ACTCTTTGCT ACCTATCCAG GCGGTCTCAT CACTGTTTCG CATGACCGTC GTTTCTTAAA 11940

AGAAGTCTGC TCGATCATCT ATCGCATGAC AGAACACGGT TTGAAGCTAG TTAATTTAGA 12000

AGATTTATAA ATTTGCAACA TAGCAAAAAT CCAGAGACGA CCTCTGGATT CTTTTACATC 12060

TGTTTTAAAC GTTCAATCCG TTCTGAGATA GGTGGGTGGG TATAAAAGAG TTTTTGGAAC 12120

CCCCCACCTT TCTTAGGATC ATTGATATAA AGGGCACTGC TAGCATCATC GACGTGGCGA 12180

CTCATAGGTT TGCTATTGTC CAACTTATCT AGGGCATTAA TCATTCCCTG GGGATTGCGA 12240

GTCAGCTCGA CACTAGATGC ATCTGCCAGA AATTCCCTCT GACGAGAAAT AGCGAGCTGA 12300

ACCAAGGTTG CAGCGAGAGG TGCCAGTACA ATAGCTAGTA GGGAAACCAC TAGCATAATG 12360

ATTTCAAGAC CATTTCCATC TCGGTCATCA TCACTTCGTC TGCGACCTGC TCCACCCCAC 12420

CACATCATAC GACCTGCCAT ACTAGAAAGC ATGGTGATAG CACTAGCAAG GGCAACTGCA 12480

ATAGTCGAAA TACGGATATC ATAATTACGA ATATGACTGA CTTCATGTCC CATAACAGCT 12540

TCTAGTTCTT CACGATTCAT GATAGCTAGT AGACCTGAAG TCGCAGCAAC AGCCGCATTT 12600

TGAGGATTAG AACCTGTCGC AAAGGCATTT AAGGCTGGAT CATCAATGAT GAAAACACGG 12660

GGCATAGGAA TCTGAGCGAC CAGAGCCATA TCTTCCACTA CATGGTAGAG GTCTGGTGCC 12720

GTTTGCTCAT CCACCTCACG CGCTCCATTC ATGGACATGA CAATCTCTGT CGATTGAAAA 12780

ATCATAGACA AAGCGTAGAT AAAGCCGATA ATCAGTGCAA TAACCAAACC ACCAAGTCCA 12840

GATCTTATAA AGAGATAACC AACCGCATAA CCAACAAGAG CTAAGAGTAG GAAAAATACC 12900

AGCAACAAAA TCCAGGTTTT TCGTTTATTG CTTGCAATTT GATCAAACAA CATCTTAGTC 12960

ACCTAAACCG CTAAAATCAA CTTTAGGAAC CGACTTTTCC TCTTCAGGTG TTTGAAGGAA 13020

ATCTGCCGCT TTAAATCCAA ACATTCCAGC GATAATATTG CTCGGGAAAG TTTCTAATTT 13080

TACATTGTAG TTGCTGACAA CACTGTTATA GAGTTGACGA GAGTAAGAAA TTTTATTTTC 13140

TGTGTTTGTC AACTCCTCTT GCAATTTAAC AAAGTTAGCA CTAGCTTTCA AATCTGGATA 13200

GCTTTCTGCA ACTGCAAAAA TACCTGAAAC CTGACGAGTG AGGGCATCAC TGGCTTTCAT 13260

AGCTTCTGCT GGTGAAGTCG CTGCCGCCAC TTGGTTACGT AGTTCTGCCA CCTTTTCAAG 13320

GGTAGAACCT TCATATTTGG CATAACCTTT TACAGTCTCA ATCAAGTTTG GCAAGAGGTC 13380

ATTGCGACGT TTCAACTGAA CATCAATCTG ACTCCAAGCC TCCTTGGTTT GCATACGATT 13440

TTTAACCAAA CCGTTATAGC TAACAATCAC AAAAATAACA ATAAGAGCGA TAACTCCAAG 13500

AATAATCCAA GTCATAATAT AAGTCCTTTC TGCTTTTAGA TTAGTACCAG TATATCAAAT 13560

TTTCTATGAT TGTGGTAAAA TAAGATGATA CTAAAGAAGG AAATAACTAT GAAACCAAAA 13620

ACATTTTACA ACTTGCTTGC CGAGCAGAAT CTTCCACTTT CGGACCAGCA AAAAGAACAA 13680

TTTGAACGTT ATTTTGAGCT CTTGGTCGAG TGGAATGAGA AGATTAATTT GACGGCGATT 13740

ACGGACAAGG AAGAAGTTTA TCTCAAACAT TTTTACGATT CGATTGCACC CATTCTTCAA 13800

GGTTTGATTC CCAATGAAAC TATCAAACTT CTTGATATCG GGGCTGGGGC AGGATTTCCT 13860

AGTCTACCAA TGAAAATTCT CTATCCGGAG TTAGATGTGA CCATTATTGA TTCACTCAAT 13920

AAGCGCATCA ACTTCCTACA ACTCTTGGCT CAAGAACTGG ATTTGAACGG AGTTCATTTC 13980

TACCACGGAC GTGCCGAAGA TTTTGCCCAA GACAAGAACT TCCGTGCTCA ATATGATTTT 14040

GTAACAGCTC GTGCGGTTGC CCGTATGCAG GTCCTATCTG AATTGACTAT TCCCTACCTT 14100

AAGGTTGGTG GCAAACTATT AGCACTCAAG GCTAGCAATG CGCCTGAGGA ATTATTAGAA 14160

GCTAAGAATG CCCTCAATCT CCTTTTTAGT AAGGTCGAAG ACAATCTCAG TACGCCCTAC 14220

CGAATAGAGA TCCGCGCTAT ATCACAGTGG TAGAAAAGAA AAAAGAAACA CCAAATAAAT 14280

ATCCACGTAA GGCTGGTATG CCAAATAAAC GCCCACTTTA AATTTTTTAG TAAACAAATG 14340

TTTACAAAAT CAGCCTCGCT CTTTTATTTC TAGGCTCGGG AAAAAATGAT TTACAAAATC 14400

AGCCTCGCTC TTTTATTTCT AGGCTCGGGA AAAAATGATT TACAAAATCA TTTTTTTCTG 14460

CTATACTATC CTAAGCAAAG GTTTTTAATG TCATCCCGTG AGGTGACGAA GACGCAGAAA 14520

TATTTAAAAC TCTTTAAAAT CTAAATTTTA AAGAAGTCTT ACTCTGAGGG CCTATTGCTG 14580

TAAAATAATG GGCTCTTTTT TGATGCCCAA AAGTGAGGTT TATATGAAAC AAGAATCAAC 14640

TGTTGATTTG TTAC 14654

(2) INFORMATION FOR SEQ ID NO: 107:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6405 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

AGAAAAATCT GCTTTACAGA AAATAAAAAT AATAGGAGAA AATCTATGTC AGATTTGAAA 60

AAATACGAAG GTGTCATTCC AGCCTTCTAC GCATGTTATG ATGATCAAGG AGAAGTAAGC 120

CCAGAACGTA CGCGTGCCTT GGTTCAATAC TTCATTGATA AAGGTGTTCA AGGTCTTTAT 180

GTCAATGGTT CTTCTGGTGA ATGTATCTAC CAAAGCGTTG AAGATCGCAA GTTGATTTTG 240

GAAGAAGTCA TGGCGGTAGC AAAGGTAAAT TGACCATTAT TGCCCATGTT GCTTGCAATA 300

ATACTAAAGA TAGTATGGAA CTTGCTCGCC ATGCTGAAAG CTTGGGAGTA GATGCTATTG 360

CAACGATTCC ACCAATTTAT TTCCGCTTGC CAGAATACTC AGTTGCCAAA TACTGGAACG 420

ATATCAGTTC TGCAGCTCCA AACACAGACT ACGTGATTTA CAACATTCCT CAATTGGCAG 480

GGGTTGCTTT GACTCCAAGC CTTTACACAG AAATGTTGAA AAATCCTCGT GTTATCGGTG 540

TGAAGAACTC TTCTATGCCA GTTCAAGATA TCCAAACCTT TGTCAGCCTT GGTGGAGAAG 600

ACCATATCGT CTTTAATGGT CCTGATGAGC AGTTCCTAGG AGGACGCCTC ATGGGGGCTA 660

GGGCTGGTAT CGGTGGTACT TATGGTGCTA TGCCAGAACT CTTCTTGAAA CTCAATCAGT 720

TGATTGCGGA TAAGGACCTA GAAACAGCGC GTGAATTGCA GTATGCTATC AACGCAATCA 780

TTGGTAAACT CACTTCTGCT CATGGAAATA TGTACGGTGT CATCAAAGAA GTCTTGAAAA 840

TCAATGAAGG CTTGAATATT GGATCTGTTC GTTCACCATT GACACCAGTG ACTGAAGAAG 900

ATCGTCCAGT TGTAGAAGCG GCTGCTGCCT TGATTCGTGA AACCAAGGAG CGCTTCCTCT 960

AATCTAAAAG GAGGTATTTA TGACATATTA CGTTGCAATT GATATCGGTG GAACCAACAT 1020

CAAGTATGGT TTGGTTGATC AAGAGGGGCA ACTTCTTGAA TCGCATGAAA TGCCAACTGA 1080

GGCGCATAAG GGTGGACCTC ATATCTTACA AAAGACCAAA GATATCGTAG CTAGTTATTT 1140

AGAAAAAGGC CCAGTAGCAG GTGTTGCCAT ATCTTCTGCT GGGATGGTGG ATCCGGATAA 1200

GGGTGAGATT TTCTATGCTG GGCCGCAAAT CCCTAACTAC GCAGGCACCC AGTTCAAAAA 1260

GGAAATCGAA GAAAGCTTTA CTATTCCTTG TGAGATTGAA AATGATGTCA ACTGTGCAGG 1320

TCTTGCTGAG GCAGTATCTG GTTCAGGCAA GGGAGCAAGT GTGACACTTT GCTTGACCAT 1380

TGGAACCGGT ATCGGTGGTT GCTTGATTAT GGATAGGAAA GTCTTCCATG GTTTTAGCAA 1440

TTCAGCCTGT GAAGTCGGGT ATATGCATAT GCAGGATGGA GCTTTTCAAG ACTTGGCTTC 1500

TACAACAGCT TTAGTGAAAT ATGTAGCTGA AGCCCATGGA GAAGATGTTG ATCAGTGGAA 1560

TGGCCGTAGA ATTTTCAAAG AAGCCACTGA AGGAAACAAA ATCTGCATGG AAGGTATTGA 1620

CCGTATGGTT GACTATCTAG GAAAAGGTCT GGCAAATATT TGCTACGTTG CCAATCCAGA 1680

AGTGGTTATT CTTGGTGGTG GTATCATGGG GCAAGAGGCT ATCCTCAAAC CTAAGATCCG 1740

TACAGCCTTG AAAGAGGCTT TGGTACCAAG TTTAGCAGAA AAAACACGAT TAGAATTTGC 1800

CCATCACCAA AATACAGCAG GGATGTTGGG TGCATATTAT CATTTTAAGA CAAAACAATC 1860

CTAGTTTGGC TCAGCCAAAC TAGGATTTTC TTACACGTTT TTGTCTACGA TAGCCGTTGA 1920

GTTTTTTATT TTCCCAGTAG CTATTAAAGA TTTTTTCCTT GCTTTCGCGA TTGATTTCCA 1980

AAAAGTAGGC ATAAATCAAA TCGATAAAGA AGAGCATAGG AAGTTGAGCG GATATTCGTT 2040

GGATATAGGA GGGTTGGCTG TGGGTGGCTA CAAGAACAGT CTCTGTATAG GTCTGGCTAT 2100

CTTTATTGGG AACACTTGTA AAGAGTACAG TCTTTGCCCC CATCTCCTTA GCATCTAATA 2160

GACTATCTAA AATAGAAGGA GTTGAGCCTG AAAGTGAGAA GCCCAGTACT AGACAATTTT 2220

CATCCATGAT GCTGGTTGTC CAGGCAAAGC CGTCTTGGTC TGTCAAAGCT TCGCAGACCA 2280

CACCTAGTCG CATAAAACGT AATTTCATTT CACGGGCGAC GAGGCCAGAA CTCCCTGTTC 2340

CAAAGAAGTA GATACGCTCA GCATCTTCGA TTAGCTGGGC AATTCGTTCT AGTTGGATTT 2400

CGTCAATCAA GTCTTGTGTT TGTTCCCTCA TATTGCTATA ACTTCTGAGG ACTCGTTTGG 2460

TCAGTGGACT GTGCTTGGAG ACTTGGTTGG CTTGATTTTC TGCCTGATGT TGGTATTGGA 2520

AAATAAATTC TCGGTAGCCA GTAAAGCCAC ACTTTTTAGC AAAGCGGGTC AAAGCAGCTT 2580

GAGAAATATG TAATTTTTGG GTGACTTGTT GAGAAGATAA ATCATCTGTA ATCGTTTCAG 2640

CTTGCAAAAA ATAGCGAGCG ATTTCTTGTT CTAGGTCTGT CATTTCTTCA AAATGTGAAT 2700

CAATGATAGT TGCGATATCT GGTTTGTCCA TAGGGAAAGC TCCTTTACAT GAGTCATACT 2760

GGAAGACTAG ATCAGAGAAT AGTCACACTT CATTATAACA CATAATATAA GGATAGATAA 2820

ATAAAAACGC ATCTCTGTTT TAAAAACGAA AAAATCGAAA AAGCTTCTCT CTTTTCCATA 2880

ATTTTCTACT CAAATTGTGG TACAATTAAG AGTAAGATTT TAAGTTAGAA ATGAGACTGA 2940

TTTGTATGAG AAAATTTAAC AGCCATTCGA TTCCGATTCG GCTTAATTTA TTGTTTTCAA 3000

TCGTCATTTT ACTCTTTATG ACCATTATTG GTCGTTTGTT GTATATGCAG GTTTTGAACA 3060

AGGATTTTTA CGAAAAAAAG CTAGCTTCAG CTAGTCAGAC CAAGATTACA AGCAGTTCAG 3120

CCCGTGGGGA AATTTATGAT GCTAGTGGAA AACCTTTGGT AGAAAATACG TTAAAGCAGG 3180

TTGTTTCCTT TACGCGTAGC AATAAAATGA CGGCTACAGA CTTAAAAGAA ACAGCTAAAA 3240

AGTTACTGAC TTATGTGAGC ATCAGTTCTC CAAATTTGAC AGAACGCCAG CTGGCGGATT 3300

ACTATTTGGC TGATCCTGAA ATCTATAAAA AAATAGTGGA AGCTCTCCCA AGTGAGAAAC 3360

GCTTGGATTC AGATGGCAAT CGTCTATCCG AATCAGAACT GTATAACAAT GCGGTCGATA 3420

GTGTACAAAC GAGTCAACTA AACTATACAG AGGATGAAAA GAAAGAAATC TATCTTTTTA 3480

GTCAGTTAAA TGCTGTTGGA AACTTTGCGA CAGGAACCAT TGCGACAGAT CCTCTAAATG 3540

ATTCTCAGGT GGCTGTTATT GCCTCTATTT CAAAGGAGAT GCCTGGCATT AGTATTTCTA 3600

CTTCTTGGGA TAGAAAGGTT TTGGAAACTT CCCTTTCTTC TATAGTTGGG AGTGTATCCA 3660

GTGAAAAAGC TGGTCTCCCA GCGGAAGAAG CAGAAGCCTA TCTTAAAAAA GGCTATTCTC 3720

TAAATGACCG TGTAGGAACC TCCTATTTGG AAAAGCAATA TGAAGAGACC TTACAAGGAA 3780

AACGCTCGGT AAAAGAAATC CATCTGGATA AATATGGCAA TATGGAAAGC GTGGATACAA 3840

TTGAGGAAGG TAGTAAGGGA AACAATATCA AACTGACCAT TGATTTGGCT TTCCAAGATA 3900

GCGTGGATGC TTTACTGAAA AGTTATTTCA ATTCTGAGCT AGAAAATGGT GGAGCCAAGT 3960

ATTCTGAAGG TGTCTATGCA GTCGCCCTTA ACCCAAAAAC AGGTGCGGTT TTGTCTATGT 4020

CAGGGATTAA ACATGACTTG AAAACGGGAG AGTTGACGCC TGATTCCTTG GGAACGGTAA 4080

CCAATGTCTT TGTTCCAGGT TCGGTTGTCA AGGCGGCGAC CATCAGCTCA GGTTGGGAAA 4140

ATGGAGTCTT GTCAGGAAAC CAGACCTTGA CAGACCAGTC CATTGTCTTC CAAGGTTCAG 4200

CTCCCATCAA TTCTTGGTAT ACTCAGGCTT ACGGTTCATT CCCTATCACA GCGGTCCAAG 4260

CTCTGGAGTA TTCATCAAAT ACCTATATGG TCCAAACAGC CTTAGGTCTT ATGGGGCAAA 4320

CCTATCAACC CAATATGTTT GTCGGCACCA GCAATCTAGA GTCTGCTATG GAGAAACTGC 4380

GTTCAACCTT TGGCGAATAT GGCTTGGGTA CTGCGACAGG AATTGACCTA CCAGATGAAT 4440

CTACTGGATT TGTTCCCAAA GAGTATAGCT TTGCTAATTA CATTACTAAT GCCTTTGGGC 4500

AGTTTGATAA CTATACGCCG ATGCAGTTGG CTCAGTATGT AGCAACTATT GCAAATAATG 4560

GTGTTCGTGT GGCTCCTCGT ATTGTTGAAG GCATTTATGG TAATAATGAT AAGGGAGGAC 4620

TGGGTGACTT GATTCAGCAA CTGCAACCGA CAGAGATGAA TAAGGTCAAT ATATCCGACT 4680

CCGATATGAG CATCTTGCAC CAAGGTTTTT ATCAGGTTGC CCATGGTACT AGTGGATTGA 4740

CAACTGGACG TGCCTTTTCA AATGGTGCCT TGGTATCCAT TAGCGGAAAA ACAGGTACAG 4800

CCGAAAGCTA TGTGGCAGAT GGTCAGCAAG CAACCAATAC CAATGCGGTG GCCTATGCCC 4860

CATCTGATAA TCCCCAAATC GCTGTCGCAG TGGTCTTTCC TCATAATACC AATCTAACAA 4920

ATGGTGTAGG ACCTTCCATT GCGCGTGACA TTATCAATCT GTATCAAAAA TACCATCCAA 4980

TGAATTAGAA AGGAAATTAT GCTTTATCCA ACACCTATTG CCAAGTTGAT TGACAGTTAT 5040

TCTAAGTTAC CAGGTATCGG GATTAAGACG GCTACGCGTC TGGCCTTTTA TACGATTGGG 5100

ATGTCTGCTG ATGATGTCAA TGAATTTGCA AAAAATCTCC TTTCTGCTAA GAGAGAATTG 5160

ACATATTGTT CTATTTGTGG ACGTTTGACA GACGACGATC CTTGTTCTAT CTGTACTGAT 5220

CCGACTCGTG ACCAGACAAC AATTTTAGTT CTTGAGGATA GTAGAGATGT GGCAGCCATG 5280

GAAAATATCC AAGAATACCA TGGACTCTAT CATGTCCTTC ATGGCCTCAT TTCTCCTATG 5340

AATGGTATCA GTCCGGACGA TATCAATCTC AAGAGCCTTA TGACTCGTCT TATGGATAGT 5400

GAGGTTTCAG AAGTGATTGT GGCGACTAAT GCTACAGCGG ATGGTGAAGC GACTTCCATG 5460

TATCTTTCAC GTTTGCTCAA GCCGGCTGGT ATCAAGGTTA CGCGTCTAGC ACGAGGTCTC 5520

GCTGTGGGAG CGGACATTGA GTATGCGGAC GAAGTGACAC TCTTACGAGC CATTGAAAAT 5580

CGGACAGAGT TGTAAGTGTA GGCAAATTTA CGAACTCCAT TCATTTATAA AAAATCAAAG 5640

AGGCTGAAAA TCGTTCCTAT CGGCCTCTTT TTGTATAGTG TGATGAGTAG GCTCAGGTTC 5700

AAGTTTTAAA AAACCAAGCA AATATGATAT ACTAAAGAGC GAGTATTCTA GTAGAATTAG 5760

GACAAATAAT ATGAAACAAA CGATTATTCT TTTATATGGT GGACGGAGTG CGGAACGCGA 5820

AGTCTCTGTC CTTTCAGCTG AGAGTGTCAT GCGTGCGGTC GATTACGACC GTTTCACAGT 5880

CAAGACTTTC TTTATCAGTC AGTCAGGTGA CTTTATCAAA ACACAGGAAT TTAGTCATGC 5940

TCCGGGGCAA GAAGACCGTC TCATGACCAA TGAAACCATT GATTGGGATA AGAAAGTTGC 6000

ACCAAGTGCT ATCTACGAAG AAGGTGCAGT GGTCTTTCCA GTCCTTCACG GGCCAATGGG 6060

AGAAGATGGC TCTGTTCAAG GATTCTTGGA AGTTTTGAAA ATGCCTTACG TTGGTTGCAA 6120

CATTTTGTCA TCAAGTCTTG CCATGGATAA AATCACGACT AAGCGTGTTC TGGAATCTGC 6180

TGGTATTGCC CAAGTTCCTT ATGTGGCTAT CGTTGAAGGC GATGATGTGA CTGCTAAAAT 6240

CGCTGAAGTG GAAGAAAAAT TGGCTTATCC AGTCTTCACT AAGCCGTCAA ACATGGGGTC 6300

TAGTGTCGGT ATTTCTAAGT CTGAAAACCA AGAAGAACTC CGTCAAGCCT TAAAACTTGC 6360

CTTCCGATAT GACAGCCGTG TCTTGGTTGA GCAAGGAGTG AATGC 6405

(2) INFORMATION FOR SEQ ID NO: 108:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11309 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

CGAGCTCGGG TACCGGGATT TTAAGGAGTT TGATATGTAT AACCTATTAT TAACCATTTT 60

ATTAGTATTA TCTGTTGTGA TTGTGATTGC AATTTTCATG CAACCAACCA AAAACCAATC 120

CAGCAATGTA TTTGATGCCA GTTCAGGTGA TTTGTTTGAA CGCAGTAAAG CTCGCGGTTT 180

TGAAGCTGTA ATGCAGCGTT TGACAGGGAT TTTAGTCTTT TTCTGGCTAG CCATTGCCTT 240

AGCATTGACG GTATTATCAA GTAGATAAGA AAATAATGGG CAGGACTAGG TCTTTGCCTC 300

TTTTTATTTT TAAAGGATGT TTGAGAAGGT TTTACAGTAA AAGAAAATTA AAAAATCTAG 360

AAAGAAAATA TGAAAGATAG AATAAAAGAA TATTTACAAG ACAAGGGAAA GGTGACTGTT 420

AATGATTTGG CTCAGGCTTT GGGAAAAGAC AGTTCCAAGG ATTTTCGTGA GTTGATTAAA 480

ACCTTGTCCT TAATGGAAAG AAAGCACCAA ATTCGTTTTG AAGAAGATGG TAGTCTGACA 540

TTAGAAATTA AGAAAAAACA TGAGATTACC CTCAAGGGGA TTTTTCATGC CCATAAAAAT 600

GGCTTTGGCT TTGTTAGTCT GGAAGGCGAG GAGGACGACC TTTTTGTAGG GAAAAATGAT 660

GTCAACTATG CTATTGATGG TGATACCGTC GAGGTAGTGA TTAAGAAAGT CGCTGACCGC 720

AATAAGGGAA CAGCAGCAGA AGCCAAAATT ATTGATATCC TAGAACACAG TTTGACAACA 780

GTTGTCGGGC AAATCGTTCT GGATCAGGAA AAACCTAAGT ATGCTGGCTA TATTCGTTCA 840

AAAAATCAGA AAATCAGTCA ACCGATTTAT GTTAAGAAAC CAGCCCTAAA ATTAGAAGGA 900

ACAGAAGTTC TCAAAGTCTT TATCGATAAA TACCCAAGCA AGAAACATGA TTTCTTTGTC 960

GCGAGTGTTC TCGATGTAGT GGGACACTCA ACGGATGTCG GAATTGATGT TCTTGAGGTC 1020

TTGGAATCAA TGGACATTGT ATCCGAGTTT CCAGAAGCTG TTGTTAAGGA AGCAGAAAGT 1080

GTGCCTGATG CTCCGTCTCA AAAGGATATG GAAGGTCGTC TGGATCTAAG AGATGAAATT 1140

ACCTTTACCA TTGACGGTGC GGATGCCAAG GACTTGGACG ATGCAGTGCA TATCAAGGCT 1200

CTGAAAAATG GCAATCTGGA GTTTGGGGTT CACATCGCAG ATGTTTCTTA TTATGTGACC 1260

GAGGGGTCTG CCCTTGACAA GGAAGCCCTT AACCGTGCGA CTTCTGTTTA CGTGACAGAC 1320

CGAGTGGTGC CAATGCTTCC AGAACGACTA TCAAATGGCA TCTGCTCTCT CAATCCCCAA 1380

GTTGACCGCC TGACCCAGTC TGCTATTATG GAGATTGATA AACATGGTCG TGTGGTCAAC 1440

TATACCATTA CACAAACAGT TATCAAGACC AGTTTTCGTA TGACCTATAG CGATGTCAAT 1500

GATATCCTAG CTGGCGATGA AGAAAAGAGA AAAGAATATC ATAAAATTGT ATCAAGTATC 1560

GAACTCATGG CCAAGCTTCA TGAAACTTTA GAAAACATGC GTGTGAAACG TGGAGCTCTC 1620

AATTTTGATA CCAATGAAGC GAAGATTTTA GTGGATAAAC AAGGTAAGCC TGTTGATATC 1680

GTTCTTCGGC AGCGTGGTAT TGCCGAGCGG ATGATTGAGT CTTTTATGTT GATGGCTAAT 1740

GAAACAGTTG CCGAACATTT CAGCAAGTTG GATTTGCCTT TTATCTATCG AATTCACGAG 1800

GAGCCTAAGG CTGAAAAGGT TCAGAAGTTT ATTGATTATG CTTCGAGTTT TGGCTTGCGC 1860

ATTTATGGAA CTGCCAGTGA GATTAGTCAG GAGGCACTTC AAGACATCAT GCGTGCTGTT 1920

GAGGGAGAAC CTTATGCAGA TGTATTGTCC ATGATGCTTC TTCGCTCTAT GCAGCAGGCT 1980

CGTTATTCGG AGCACAATCA CGGCCACTAT GGACTAGCTG CTGACTATTA TACTCACTTT 2040

ACCAGTCCAA TTCGTCGTTA TCCAGACCTT CTTGTTCACC GTATGATTCG GGATTACGGC 2100

CGTTCTAAGG AAATAGCAGA GCATTTTGAA CAAGTGATTC CAGAGATTGC GACCCAGTCT 2160

TCCAACCGTG AACGTCGTGC CATAGAAGCT GAGCGTGAAG TCGAAGCCAT GAAAAAGGCT 2220

GAGTATATGG AAGAATACGT GGGTGAAGAG TATGATGCAG TTGTATCAAG TATTGTCAAA 2280

TTCGGTCTCT TTGTCGAATT GCCAAACACA GTTGAAGGCT TGATTCACAT CACTAATCTG 2340

CCTGAATTTT ATCATTTCAA TGAGCGTGAT TTGACTCTTC GTGGAGAAAA ATCAGGTATC 2400

ACTTTCCGAG TGGGTCAGCA GATCCGTATC CGTGTTGAAA GAGCGGATAA AATGACTGGA 2460

GAGATTGATT TTTCATTCGT ACCTAGTGAG TTTGATGTGA TTGAAAAAGG CTTGAAACAG 2520

TCTAGTCGTA GTGGCAGAGG GCGTGATTCA AATCGTCGTT CGGATAAGAA GGAAGACAAG 2580

AGAAAATCAG GACGCTCAAA TGATAAGCGT AAGCATTCAC AAAAAGACAA GAAGAAAAAA 2640

GGAAAGAAAC CTTTTTACAA GGAAGTAGCT AAGAAAGGAG CCAAGCATGG CAAAGGGCGA 2700

GGGAAAGGTC GTCGCACAAA ATAAAAAGGC ACGCCACGAC TATACAATCG TAGATACGCT 2760

AGAGGCAGGG ATGGTCCTGA CTGGAACTGA AATCAAGAGT GTACGAGCTG CTCGAATTAA 2820

TCTCAAGGAT GGCTTTGCTC AAGTGAAAAA TGGAGAAGTT TGGCTGAGCA ATGTTCATAT 2880

CGCGCCTTAC GAAGAGGGCA ATATCTGGAA CCAGGAACCA GAACGTCGTC GTAAACTCCT 2940

GCTCCATAAA AAGCAAATTC AAAAATTGGA ACAAGAGATC AAAGGGACAG GAATGACCTT 3000

AGTTCCCCTT AAGGTCTATA TAAAAGATGG CTACGCTAAG CTTCTTTTAG GACTTGCCAA 3060

AGGGAAGCAT GACTATGACA AACGGGAGTC TATCAAACGT CGTGAGCAAA ATCGAGATAT 3120

CGCGCGTGTG ATGAAAGCTG TTAATCAGCG ATAAAAAGAG GAATTGAAAA TGGAAAAATT 3180

AGTTGCCTAT AAACGCATGC CTTTGTGGAA TAAACAAACA ATGCCTGAAG CTGTTCAGCA 3240

AAAGCACAAT ACAAAAGTTG GGACTTGGGG GAAAATTACT GTCTTGAAGG GAGCTCTCAA 3300

GTTTATTGAA TTGACAGAAG AAGGGGAAGT TCTAGCTGAA CACCTCTTTG AAGCAGGGGC 3360

AGACAATCCA ATGGCCCAAC CTCAAGCCTG GCACCGAGTG GAAGCTGCCA CAGATGATGT 3420

GGAATGGTAC TTGGAATTTT ATTGTAAACC TGAGGATTAT TTTGCTAAAA AATACAATAC 3480

CAATCCTGTT CATTCAGAGG TCCTAGAGGC CATGCAGACA GTGAAACAAG GGAAAGCTTT 3540

GGATTTGGGT TGTGGTCAGG GGCGTAATTC TCTTTTTCTA GCCCAGCAAG ATTTTGATGT 3600

GACGGCTGTA GATCAAAATG GACTAGCTCT TGAAATCTTG CAAAGCATTG TGGAGCAGGA 3660

AGATTTGGAC ATGCCTGTTG GCCTTTACGA TATCAATTCA GCTAGCATTG AACAAGAATA 3720

TGATTTTATC GTTTCAACAG TTGTTCTCAT GTTTCTACAA GCCGACCGCA TTCCAGCTAT 3780

TATTCAAAAT ATGCAGGAGA AAACCAGTGT TGGTGGTTAC AACCTTATCG TTTGTGCCAT 3840

GGACACGGAG GATTATCCTT GCTCGGTTAA CTTCCCATTC ACCTTTAAAG AAGGAGAACT 3900

GGCAGACTAT TACAAGGATT GGGAATTGGT TAAGTACAAT GAAAATCCAG GCCATTTGCA 3960

CCGTCGCGAT GAGAATGGCA ATCGTATTCA ACTACGCTTT GCGACCTTAC TAGCTAAGAA 4020

AATCAAGTAA ACACACATGA AGATTAGGAA TTTTCCTGAT CTTTTTTCTT TTTTACGAAT 4080

GATATAGAAA AGGAGGGAAT TCATGTTTGT TGCGAGAGAT GCTAGGGGAG AATTGGTAAA 4140

TGTGTTAGAG GATAAACTTG AGAAGCAAGC ATACACCTGC CCAGCTTGTG GAGGCCAGCT 4200

CCATTTGCGT CAAGGACCAA GTGTACGGAC GCATTTTGCC CATAAATCCT TAAAAGACTG 4260

TGATTTTTTC TTTGAAAATG AAAGTCCAGA ACACCTGGCC AATAAGGAAT CCCTCTATCA 4320

CTGGTTGAAA AAAGAGACAA AGGTTCAATT AGAGTACCCG CTTTCAGAAC TTAAACAGAT 4380

TGCGGATGTA TTTGTAAATG GCAATCTAGC TCTAGAAGTT CAGTGTAGTC CCTTGCCTCA 4440

GAAAGTCCTT AAAGAGCGAA GTGAGGGCTA TCGTAGTCAG GGTTACCAAG TACTGTGGTT 4500

GCTGGGTCAA AAACTGTGGC TCAAGGAGCG TTTGACTCGT CTACAGCAAG GTTTTCTTTA 4560

TTTCAGTCAA AACATGGGCT TTTATGTTTG GGAATTAGAC AAGGAAAAAC AAGTTTTAAG 4620

ACTCAAATAC CTGATTTACC AGGATCTCCG CGGTAAACTC CATTATCAAA TCAAGGAATT 4680

TTCCTATGGT CAAGGTAGTT TATTGGAAAT ATTGCGTCTT CCCTATAAGA GACAAAAAAT 4740

ATCTCATTTT ACAGTTTCTG AGGACAAGGA CATCTGTCGC TATATCCGGC AACAACTTTA 4800

TTATCAAAAT CTCTTTTGGA TGAAAGAACA AGCAGAAGCC TATCAAAAGG GAGAAAATAT 4860

CCTGACTTAT GGACTGAAAG AATGGTATCC ACAAATTCGA CCAATAGTGG GCAAATTTTT 4920

CCAGATTGAA CAAGACTTGA CTAGCTATTA TCAGCACTTT TATACCTATT ACCAAAAAAA 4980

TCCTCAAAAT GATTGGCAAA AGCTTTATCC ACCAGCCTTT TATCAGCAAT ATTTCTTGAA 5040

AAATATGGTA GAATAGAAAG GATGGAGGAA TCTAATGGTA TTACAAAGAA ATGAAATAAA 5100

TGAAAAAGAT ACATGGGATC TATCAACGAT CTACCCAACT GACCAGGCTT GGGAAGAAGC 5160

CTTAAAAGAT TTAACAGAAC AATTGGAGAC AGTAGCCCAG TATGAAGGCC ATCTCTTGGA 5220

TAGTGCGGAT AACCTACTAG AAATCACTGA ATTTTCTCTT GAAATGGAAC GCCAGATAGA 5280

GAAGCTTTAC GCTTATGCTC ATATGAAGAA TGACCAGGAT ACACGTGAAG CTAAGTATCA 5340

AGAGTACTAT GCCAAGGCCA TGACACTCTA CAGCCAGTTA GACCAAGCCT TTTCATTCTA 5400

TGAGCCTGAA TTTATGGAGA TTAGCGAAAA GCAGTATGCT GACTTTTTAG AAGCTCAACC 5460

AAAGCTGCAG GTTTATCAAC ACTATTTTGA CAAGCTTTTG CAAGGCAAGG ATCACGTTCT 5520

TTCACAACGT GAAGAAGAAT TATTGGCTGG AGCTGGAGAA ATCTTTGGTT CAGCAAGTGA 5580

AACCTTCGCT ATCTTGGACA ATGCGGATAT TGTGTTCCCT TATGTCCTAG ACGATGATGG 5640

TAAAGAAGTT CAGCTATCTC ATGGGACTTA CACACGTTTG ATGGAGTCTA AAAAACGTGA 5700

GGTTCGCCGT GGTGCCTATC AAGCTCTTTA TGCGACTTAC GAACAATTCC AACACACCTA 5760

TGCCAAAACC TTGCAAACCA ATGTTAAGGT GCAAAATTAC CGTGCTAAAG TTCGTAACTA 5820

CAAGAGTGCT CGTCATGCAG CCCTCGCAGC GAATTTTGTT CCAGAAAGTG TTTATGACAA 5880

TTTGGTAGCA GCAGTTCGCA AGCATTTGCC ACTCTTACAT CGCTATCTTG AGCTTCGTTC 5940

AAAAATCTTG GGGATTTCAG ATCTCAAGAT GTACGATGTC TACACACCGC TTTCATCTGT 6000

TGAATACAGT TTTACCTACC AAGAAGCCTT GAAAAAAGCA GAAGATGCTT TGGCAGTCTT 6060

GGGTGAGGAT TACTTGAGCC GTGTTAAACG TGCCTTCAGC GAGCGTTGGA TTGATGTTTA 6120

CGAAAATCAA GGCAAGCGTT CAGGTGCCTA CTCTGGTGGT TCTTATGATA CCAATGCCTT 6180

TATGCTTCTC AACTGGCAAG ACAATCTGGA CAATCTCTTT ACTCTTGTTC ATGAAACAGG 6240

TCACAGTATG CATTCAAGCT ATACTCGTGA AACTCAGCCT TATGTTTACG GGGATTACTC 6300

TATCTTTTTG GCTGAGATTG CCTCAACTAC CAATGAAAAT ATCTTGACGG AGAAATTATT 6360

GGAAGAAGTG GAAGACGACG CAACACGCTT TGCTATTCTC AATAACTTCC TAGATGGTTT 6420

CCGTGGAACA GTTTTCCGCC AAACTCAATT TGCTGAGTTT GAACACGCCA TTCACCAAGC 6480

AGATCAAAAT GGGGAGGTCT TGACAAGCGA TTTCCTAAAT AAACTCTACG CAGACTTGAA 6540

CCAAGAGTAT TATGGTTTGA GTAAGGAAGA CAATCCTGAA ATCCAATACG AGTGGGCTCG 6600

CATTCCACAC TTCTACTATA ACTACTATGT ATATCAATAT TCAACTGGCT TTGCGGCCGC 6660

CTCAGCCTTG GCTGAAAAAA TTGTCCATGG TAGTCAAGAA GACCGTGACC GCTATATCGA 6720

CTACCTCAAG GCAGGTAAGT CGGACTATCC ACTTAATGTC ATGAGAAAAG CTGGTGTTGA 6780

TATGGAGAAG GAAGACTACC TCAACGATGC CTTTGCAGTC TTTGAACGCC GTTTAAATGA 6840

GTTTGAAGCC CTTGTTGAAA AATTAGGATT GGCATAAAAT GGTTGAATCG TATAGTAAGA 6900

ATGCTAACCA TAACATGCGT CGTCCTGTCG TCAAAGAAGA AATTGTAGAC TTGATGCGTC 6960

AGCGTCAAAA GCAGGTCACA GGTTTCTTGA AAGAATTGGA AGACTTTGCC CGCAAGGAAA 7020

ATATTCCTAT TATTCCCCAT GAAACGGTTG CTTATTTCCG TTTTCTTATG GAAACCATGC 7080

AGCCTAAAAA TATTCTGGAA ATTGGGACGG CTATCGGTTT TTCAGCTCTC TTGATGGCTG 7140

AACATGCGCC AAATGCTAAG ATTACAACTA TTGATCGTAA TCCAGAAATG ATTGGTTTTG 7200

CCAAGGAAAA TTTTGCCCAG TTTGACAGTC GCAAGCAAAT CACTCTCCTA GAGGGAGATG 7260

CGGTGGATGT CTTATCTACA CTGACAGAGT CTTATGATTT CGTCTTTATG GATTCTGCCA 7320

AGTCTAAATA CATCGTCTTT CTGCCAGAAA TCCTCAAACA TTTGGAAGTT GGTGGTGTGG 7380

TTGTCTTGGA TGATATTTTT CAAGGTGGTG ATGTTGCCAA GGATATTATG GAAGTCCGTC 7440

GTGGTCAGCG AACCATTTAT CGAGGCCTTC AAAAATTATT TGATGCAACC TTAGACAATC 7500

CAGAACTCAC CGCAACATTA GTGCCTTTAG GAGATGGTAT TCTCATGCTT CGTAAAAATG 7560

TAGCAGATGT TCAACTGTCT GAAAGCGAAT GATTTTCAGA AAAATTTAAG AAAAAATAGT 7620

AAAATAGATA GAGTAACACT TATCTCAAAG GAGTAGACAT GAAGAAAAAA TTATTGGCAG 7680

GTGCCATCAC ACTATTATCA GTAGCAACTT TAGCAGCTTG TTCGAAAGGG TCAGAAGGTG 7740

CAGACCTTAT CAGCATGAAA GGGGATGTCA TTACAGAACA TCAATTTTAT GAGCAAGTGA 7800

AAAGCAACCC TTCAGCCCAA CAAGTCTTGT TAAATATGAC CATCCAAAAA GTTTTTGAAA 7860

AACAATATGG CTCAGAGCTT GATGATAAAG AGGTTGATGA TACTATTGCC GAAGAAAAAA 7920

AACAATATGG CGAAAACTAC CAACGTGTCT TGTCACAAGC AGGTATGACT CTTGAAACAC 7980

GTAAAGCTCA AATTCGTACA AGTAAATTAG TTGAGTTGGC AGTTAAGAAG GTAGCAGAAG 8040

CTGAATTGAC AGATGAAGCC TATAAGAAAG CCTTTGATGA GTACACTCCA GATGTAACGG 8100

CTCAAATCAT CCGTCTTAAT AATGAAGATA AGGCCAAAGA AGTTCTCGAA AAAGCCAAGG 8160

CAGAAGGTGC TGATTTTGCT CAATTAGCCA AAGATAATTC AACTGATGAA AAAACAAAAG 8220

AAAATGGTGG AGAAATTACC TTTGATTCTG CTTCAACAGA AGTACCTGAG CAAGTCAAAA 8280

AAGCCGCTTT CGCTTTAGAT GTGGATGGTG TTTCTGATGT GATTACAGCA ACTGGCACAC 8340

AAGCCTACAG TAGCCAATAT TACATTGTAA AACTCACTAA GAAAACAGAA AAATCATCTA 8400

ATATTGATGA CTACAAAGAA AAATTAAAAA CTGTTATCTT GACTCAAAAA CAAAATGATT 8460

CAACATTTGT TCAAAGCATT ATCGGAAAAG AATTGCAAGC AGCCAATATC AAGGTTAAGG 8520

ACCAAGCCTT CCAAAATATC TTTACCCAAT ATATCGGTGG TGGAGATTCA AGCTCAAGCA 8580

GTAGTACATC AAACGAATAG TCCAAATCAA TGAGTCAGGG AAAAAACTCG ACTTCAGGAA 8640

AAAATGAAGC AAACATTCCC ACAATAAAAC GCATAGTACA AGGTTTGTAC TGCCCCCCAA 8700

AAAGTTAGAC AATTAATTTA TCCGAAGGAT TTAGTTCTGT ATTGCACAGA GCTAAGTCCT 8760

TTTAGTTTTA TCTTAATTCT CTTATTGTTG TAATAATCAA TATAGTCTAT AATGGCTCGT 8820

TCCAATTGAT TAAGTGATTT AAATGTTTTC TCATAGCCAT AAAACATTTC GGATTTTAAA 8880

ATGCCAAAGA AAGATTCCAT CCTACCGTTG TCTTGGCTGT TGCCCTTACG TGACATGGAT 8940

GCTTGAATTC CCTTACTCTC TAGGAAGCGA TGATAAGAAT CGTGTTGATA TTGCCAGCCT 9000

TGGTCACTAT GGAGAATCGT ATTCTCGTAG TGCTTCTCTT TGAATGCCTG TTCCAACATT 9060

AACGATCAAT CAATTTAATC ATGTACCTAA GATTAGAATT GTTTATCCCA AATTTATTTG 9120

AAAGCTTCTC TAAGCTATAT CCTTGTTTTC TAAGTTCATA GATCTGAACT TTATCATCAT 9180

AAGTTAATTT CATAATAAAA ACACCCCAAA AGTTAGATTT TTTCTGTCTA ACTTTTGGGG 9240

TGTAGTTCAT GTACACCTGA TATGATGCGT TTTATAATTT TAAAGACTTT TTGACCAGCC 9300

TCATTTTTTT AACTTGATAC TCAGTGAAAA GCAAAGATTA AACTAGGAAG CTAGCTGTAG 9360

GCTGCTCAAA GAACAGCTTT GAGGTTGTAG ATAAAACTTG TGAGGTCACC AACATATATA 9420

ATGTGAAGCT GACGTGGTTT GAATAGATTT TAGAAGAGTA TGAGTCTGGA AGTTTTAATG 9480

GATAATGCAA GATTCCATAG AATGGGTAAG CTAGAGTTCT TATGTGAAGA GTTTGGGCAT 9540

AAACTTTTAC CTTTTCCTCC CTACTCATCT TAGTATAGAA AAGTGAATCT GAAATAGTAC 9600

ATAACTGCTT CTAAAACATT CTTATAAATT GATTTAAATT CTCAAATCAT ATTATTCAGT 9660

TCTTATTTCA TTTTGTTCTA CAATCCTGTT GAGAAGACAC GTGTTCATAT CAAAAAGGTA 9720

TTGGCAAGTT GCAATACCTT TTTACGAGGC TCTGTTGTCT TATTTTTGTT TCAACTGACT 9780

ATATCTCCTA TGGTTCTAGT TCAGAAGGCT AGGCTATAAT TATGATTGAT AAGAAGTATC 9840

ATTCCAAGTA TTGGGAGTGA ATGTTTCAAA ATCATGGGTT TCTATAATGG TCAGGCTGGC 9900

ATTTGCTAGA CCGCCATCTT TACGAAGAAG TGGTTCTTTA TAGCCTAGGA GAGTACGAAG 9960

ACTGGCAGTA AGATTGGCGC CGTGTCCGAC AATTAGAATA CGTTCAGCTG GACTATCTTT 10020

TAATGATTTG ATAAATTGGA TGGTCCGTTG AGTTGTACTA TAGAGGGATT CGGCTCCGAA 10080

CATTCGAGTG TCAAATTGAG CAAGATTTGA ACGAAAAGCC TGGATTTGTT GCGGGTAAAT 10140

AGCTTCCAAG GTTGCAATTT TCAAACCTTC TAACTTCCCA AGTTGCCATT CACGGAGATT 10200

AGGAACGATT TCTAAAGAAC AGGGGGTATA GAGTTGACTT TGGATAATCT CAGCAGATTT 10260

GACCGCTCGA GGTAAATCAC TTGAATAAAT CTGATCAAAA GGAATTTCCT TGAGATACTG 10320

800

ACCAAGTCGT TTTAGGGTTT CAATGGATTC AGGAAGAAGA GGAGAATCAC CACTAGCACC 10380

TTGAAAACGA CCTTCTTGGT TCCAGAGGGT ACGACCGTGG CGGACAAAGT AGAGTTTCAT 10440

TACTTGATGT CCTCCAAAAT ATCTACAAAG TCTGCCTTTA CAAAGCTAGC CAAGTCTTGT 10500

GGCGCGACGA TAATGCTGTG TCCGACTTCG CCTGCAGAGA CAATCATTTG ATCCAAATCT 10560

AGAGCAATTT TATCGATAAA AATGGGATAA TTGTGTTTCT GACGAATTCC GACAGGATTA 10620

TTGGCTCCAT GAATGTAACC AGTTGTTTTT TCTAAGTCCT TTTGTGGAAT CATGCTCACT 10680

TTTTTATTGC CAGAAATTTT AGCTAGTTTC TTTTCAGACA AGTGCTGAGT GATAGGGACA 10740

ATTCCGATAA TCGGTCCGGT CTTGTCTCCC AAAAGCGCCA AGGTTTTGAA AATCTGATCT 10800

CGTTCATAAC CTTGAGGAAG CTCTCCTTCT AGGGCATTGA TTTGAATCCC CTGATGAGGG 10860

ATAGCTGCTT TAGATAGGAT TTGTTCCACC AATGTTTTTT TGATTTTAAC TTTTTTTGCC 10920

ATTATTTATA TTTATCCTCC AATTGACTCA TCCAAATACC AAGCCAGATT CCCAGCGCAA 10980

AGAAGAAGGC GATGATGACA TAACCGACAA GTGAAAGTCC TGTGTATTGG ATACTTTCAG 11040

CGTTTCCTGC ATTTGGAATT AAGATCAAAA GGGTACTTGA TAGGACGATA CCGATGATGA 11100

AATGATAGAC GAACTGTTTA CGGAGTTCTT CTAGTTCTCC GTCCGTCCAA GCGTAGGCCA 11160

CTTCTTCTTT CTTGCCTTTA CCTTTGGACA TCTTGTAAAG AGGTGGGAGG GCAATATAGA 11220

CATGACCTGC CTCGACTAGC GGACGCATGT AACGGTAGAA AAATGTCAAG AGCAAGGTCT 11280

GGATATGGGC ACCGTCGGTA TCCGCATCG 11309

(2) INFORMATION FOR SEQ ID NO: 109:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5548 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

CCATAGTCTA ACAAGTCTTT GTAAAGGTTT ATCCCTGATT CATGTAAAGA TTGTGTAAAG 60

AATCAAAAAA AGCCACTTTT GAAAAATGGC TGCTCCTAAA AATAGCTTTA AAAATTATTA 120

GTCCTGTGCG AAAGATTGGT TAGGAAGAAA AATCGTGAAG CAACTGCCTC TGCCAAGCTG 180

ACTCGTCACC GTGACTTGGC CACCTAATAA TTGACTGAGT TCTTTGACAA TGGCAAGGCC 240

AAGACCAGTG CCACCAGTTT GTCTGCTTCG ACCTTTATTA ACTCGGTAAA AACGTTCAAA 300

AATACGATCC TGCTCTAATT GACTAATACC AATCCCTGTA TCTGATACAG AAATCTTAAT 360

GCCTTCGTTC ACCTTTTGGG TCTTGACCTC AATTTTTCCC CCTTGTTCAG TGTAACGGAT 420

GGCATTGGAT AAAAGATTGA GTAAGATTTG GGAAAGTAAT TGACTATCTG ATACGAGGGT 480

GACATCATCT GGCACCTGCA CCTTTAGCTG TAAATCCTTC TTCTTGAGCT GAGGTTGCAA 540

GCTTTGAGTC AAATCCTGTA CAAATTCTGC CAAAGAAAGG GTCGTCCATT GTATAGGCAT 600

TTGTTGAGCC TTAGATAAGG TAAGAAGATG CTCAACAATA TGCTCAAGAC GCAAACTTTC 660

TTTGTAAATA ATGTCTAGAA AGTCATCCTT GAGCGCTTCT TCTTCAGCTG ACATCCCCTT 720

AATGGTTTCA GCAAAGCCCT TAATCGAAGT AACTGGTGTC CTCAATTCAT GGGAGGCATT 780

TGAGACAAAG GCTAAATTTA ACTTTTCATA AGTTCTAATC GTTGTTAAAT CATATAGCAA 840

GACGAGCACA GCTTCCACAG ATTGGGTGGG GCTAAAAACG GGAACTGCTG TCACTTCTAA 900

AATCAAGTCA CCCTCATGAA ACCCACTTAC TTCTTGTTTT AACCTTGTTT TTTGATCAAA 960

GGCTTGGTGA ACTAAATTCC GAATATCCAT CCGTTTGAGG TCATCAAGTG AACTTATGTC 1020

GCCGTCCACA TCGGGAAAAT AATGAGGCAG AGAGCGACTG GATAATAACA TCTGACCTTG 1080

AGCGGAAACT AAAAACGTCC CCATGGTTAG GTGCGACAGA AGAACCTCCA TTGTTTCGGC 1140

TAGATCCTTG TATTGCTGAT CCTGTTGGGA GACTTTGGTT TTTAGGCCAG ACACATACTG 1200

AGCCAAAGAC TTTAAGTCTT CTTGCCCTTT TTCTAAAAAG TATTCACTAC TGGTCAAGAG 1260

AGGTTGGTGC AAGGTCTCAA AAGCAACTTC CCATTTCCAA AGGCAAAAGA GCCAGTAGCC 1320

ACCTAGTCCC AAAGAAAGGG CTAGAAGAAA GAGACCGATG CCTTTACTGA TCCAAGTTAA 1380

TGCCATCCCT GCAATCAGAA TGAGGCTAAC ACTTAGATTG ACTAGCCAAA ATTGAAGGTA 1440

GCGTTTCATC TATAACTCCT TGAACTTATA ACCATAACCC CGAATGGTTC GAATAAATTG 1500

AGGGGCTTTA GGATTGTCTT CAATTTTTTC CCTCAACTTA CCAATATGAA CGTCCACCAA 1560

ACGTGTTTCC TGCCCAAAGT CATACCCCCA GATACGTTCC AAAAGACGCT CTCTAGTCAG 1620

TGTCATGTTG GGATGTTTCA TAAGATAGAG CAAGAGTTCA AATTCTTTTG GGGTCAAACT 1680

CAGTAACTTA TTCGCCTTGT AGACTTCATG ACGCTCAGGG TATACTTTCA AGGTCCCAAA 1740

TAGCCAAGAA TCGTCAGCGA TATTATCTGA ATCATCTCCT TCTTGTTCTC CTTTAGTTCG 1800

CCTGAGGACA GCCTTGACAC GCGCCAGCAA TTCTCTAGGG CTAAAAGGCT TGGTCAGGTA 1860

GTCATCAGCC CCTAATTCCA AGGCCAAAAC CTTATCAAAT TCATCACTTT TCGCAGAAAC 1920

CATCATAATT GGAGTTTTGA CGCCTTTGGC TCTCAGCCGC TTACAAACTT CCATGCCATC 1980

TAATTGTGGT AACATGATAT CAAGCAAGAT AAAATCAAAG GGTTCTGTTT CTGCCAAAGC 2040

TAAGGCCTTC CGTCCATTTG TCACCAATTG AGTAGAAAAG CCTTCCTTAC TTAAATGGTA 2100

GTCAAGCAAT TTCAGAATGT GTTCTTCATC ATCCACTAAT AAGACTTGTT TTGTCATCTA 2160

TTATCTCCTA TTGGTAACAT TATAACACAA TTATCAGAAA TCCTAACATT GCTAAATCAG 2220

ATTAAATTTG CCTATCAAGA CTAGTATCTG GTCAAACGCT CAATCATCTC CTTGTGCTCT 2280

GGATAGGTCG CCAGTAGATC TACCCTTTCA AATAATTCAA AATCCTCAAA TTCAAAACCA 2340

GGAGCAACAA GACAAGAAAC CAGAGCATCA TCCTTATCAA CTGTTGATCC CCAAATAGTG 2400

CCCTTAGGAA CACAGTAGTG AAGTTGTTGC CCTTTGGATA TGTCCAGGCC TAAAGTGACT 2460

GCTTCGTAGT GACCATCTGC TGTAATCATG TGAACAGTAA GTGGGGATCC TGCATGAAAA 2520

TACCAGATTT CATCTGCTGT CAATCGGTGA AAATGTGAAG GATTCGTTTC TTCTAATAAG 2580

AAATAAATAC TGGTATAAAG CGCCCTTCCC TTACCAGCAA GGTTTATAGT GTCTGAAGCT 2640

TTTTTTGTTT GTCTAAAATA GCCACCTTCA ATATGGGGAG CTAACTCTAG AGTTCTTATC 2700

AAGTCTTCTT TATCCGTCGG AGCCAATGGG TTGAAGTAAC TCTTGTTCAA AGTGGTTTTA 2760

CGATTTCAAG AACTCCTCTC AGTTCTGAGG ACACGGTAAT GATTGATGCG ACGGAAGTAC 2820

AAATCAATCG CCCTAAAAAA AGAATTAGCG AATGATTCTG GTAAAAAAAA TGCCACGCTA 2880

TGAAGGCTCA AGCGATTGTC ACAAGTCAAG GGAGAATTGT TTCTTTGGAT ATCGCTGTGA 2940

ACTATTGTCA TGATATGAAG TTGTTCAAAA TGAGTCGCAG AAATATCGGA CAAGCTGGTA 3000

AAATCTTGGC TGACAGTGGT TATCAAGGGC TCATGAAGAT ATATCCTCAA GCACAAACTC 3060

CACGTAAATC CAGCAAACTC AAGCCACTAA CAGTTGAAGA TAAAGCCTAT AACCATGCGC 3120

TATCCAAGGA GAGAAGCAAG GTTGAGAACA TCTTTGCCAA AGTAAAAACG TTTAAAATGA 3180

TTTCAACAAC CTATCGAAAT CATCGTAAAC ACTTCGGATT ACGAATGAAT TTGATTGCTG 3240

GCATTATCAA TCATGAACTA GGATTCTAGT TTTGCAGGAA GTCTATTATT TGGTTAGGTG 3300

AATTAGTGAA GCGTTTAGGC AAGTGTCTCT GGTTACGACG TCATGGACTC TAAATCGATT 3360

ATATTTAGGG GTCATGACTA GTGAAGCAGT TAGCTAGTTC GCATATAAGC GGCTAGCGTC 3420

TAACAATTAG GAACTTTAGT TCCAATAACT TTAAGATTAC GACGTTTTAG GACATAAATC 3480

GATCATATTT ATGTCCTAAA ACTAGTGAAG CGCCTAGCCA AAGTCCGAAT AGGATTTGGC 3540

GTTAGTTACT TAGATTGCTT TGCAATCAAG TAACTTTGGC GATTTACATC TTCTCTGGCG 3600

CTTCTACTCC AAGCAAGCGA AGGGCTTCTT TGAGAACGAC TGCGGTTGCG TAGCTGAGGG 3660

CTAGACGGCT GTCGCGTTCT GGGCTTTCAT CCAAGATACG TGTATGTGCA TAGTATTTGT 3720

TAAAGGATTG AGCCAGGCTA ATTGCAAATT TAGCAATGAT AGAAGGTTCA AAGTTATCTG 3780

CCGCACGGTT GATAATACGT GGGAAGTCTT GAATGAGTTT AATGATTTCC CAGCTTTCAG 3840

TATCATTCAA GCTATAGTTG CCAGCTGTTT CTGGTTTGAA ATCGGCTTTG CGTAAGATAG 3900

ATTGGATACG AGCGTAGGCA TATTGAACGT AAGGTCCAGT TTCACCCTCG AAGGATACCA 3960

TAGCCTCTAG GTCGAAGTCG TATCCATTTG TACGGTCGGT TTTGAGGTCA TAGAATTTAA 4020

TGGCTCCAAT CCCAACAGCA TGTGCTACTT GGTCTTTGTT TTCTAGTTCA GGATTTTTAG 4080

CCTCGATTTG GACCTTGGCA CGGCTAACAG CCTCTGCAAC AGTAGGCTCT AGCAAGATGA 4140

CATTCCCTTT ACGAGTAGAG AGTTTCTTCC CTTCTTTTGT AACCAAACCA AAAGGAACGT 4200

GAGTAATGTC GTCACTCCAG TCGTAGCCCA TCTCTTGCAA GACAGCTTTG AGCTGTTTAA 4260

AGTGGGCAGA TTGTTCTTGA CCAACGACAT AGATAGATTT AGCAAATTGG TATTCGTTTT 4320

TACGGTAGAG GGCTGCAGCC AAGTCACGTG TGATATAGAG AGTTGCACCA TCAGACTTCT 4380

TGATGAGGGC TGGATGTTCA ATTCCATATT TCTCAAGATT CACAACTTGG GCACCTTCTG 4440

ATTCAAGAAG TAGTCCTTTT TCAGAAAGAA TGTCTACAAC TGCATCCATC TTATCATTGT 4500

AGAAGGCTTC TCCGTTATAG CTGTCAAATT CAACCTTCAA TTCATTGTAA AGGCGGTTAA 4560

ATTCCACTAA ACTTTCATCG CGGAACCATT GCCAAAGAGC GAGAGCTTCC TCATCTCCAT 4620

TTTCAAGTTT ACGGAACCAT TCGCGCGCTT CTTCATCCAA GCTAGGGTCA TTTTCAGCTT 4680

CAGCGTTGAT GCGGACATAG AGTTTAAGGA GTTCATCGAT TGGATGAGCT TTTACAGCTT 4740

CTTCGTCGCC CCATTTTTTG TAGGCAACAA TCAACATCCC AAATTGTTTA CCCCAGTCTC 4800

CCAAATGGTT GACCTTGACC GTTTGATAAC CGATTTTTTG GAAAATATGT GACAAGCTAT 4860

CTCCGATAAC AGTTGAACGC AGGTGGCCAA TAGAAAATGG TTTAGCGATA TTCGGACTAG 4920

ACATGTCGAT AACAACATTT TCTTGTTTAC CAATATTTTG GTCAGCATAG TGTTCTTTTT 4980

CAGTGGTAAC AGCTTGCAAT ACTTGAGCAG AAATGGCAGA TTTATCAAGG AAAAAGTTAA 5040

CGTAAGGTCC TGTTGCGACA ACTTTTTCAA AGGCTTGGCT GTTCATTTTT TCAGCCAGTT 5100

CAGCCGCAAT CATTTGTGGT GCTTTACGTT CGACTTTTGC AAGAGAAAAA GCAGGGAAAG 5160

CAATGTCTCC CATTTCTGAG TTTTTAGGGG TTTCCAGTAA CTTTAAAATA GCCTCTTGGT 5220

CCAGGCTATC AATGATGCTA GATAATTCGC TAGCAATCAA TTCTTTTGTA TTCATTAAGA 5280

GCTCCTTTTT GGACTTTTCT ACTATTTTAT CACAATTTTA AAGAAAGAAG AAAAAATTTT 5340

TGAAATCTCC TGTTTTTTTG GTATAATATG GTTATAAATA TAGTTATAAA TATGCACGCA 5400

AGAGGATTTT ATGAGAAAAA GAGATCGTCA TCAGTTAATA AAAAAAATGA TTACTGAGGA 5460

GAAATTAAGT ACACAAAAAG AAATTCAAGA TCGGTTGGAG GCGCACAATG TTTGTGTGAC 5520

GCAGACAACC TTGTCTCGTG ATTTGCGG 5548

(2) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3132 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

TACCCGGTAG TCTTAGCAGA CACATCTAGC TCTGAAGATG CTTTAAACAT CTCTGATAAA 60

GAAAAAGTAG CAGAAAATAA AGAGAAACAT GAAAATATCC ATAGTGCTAT GGAAACTTCA 120

CAGGATTTTA AAGAGAAGAA AACAGCAGTC ATTAAGGAAA AAGAAGTTGT TAGTAAAAAT 180

CCTGTGATAG ACAATAACAC TAGCAATGAA GAAGCAAAAA TCAAAGAAGA AAATTCCAAT 240

AAATCCCAAG GAGATTATAC GGACTCATTT GTGAATAAAA ACACAGAAAA TCCCAAAAAA 300

GAAGATAAAG TTGTCTATAT TGCTGAATTT AAAGATAAAG AATCTGGAGA AAAAGCAATC 360

AAGGAACTAT CCAGTCTTAA GAATACAAAA GTTTTATATA CTTATGATAG AATTTTTAAC 420

GGTAGTGCCA TAGAAACAAC TCCAGATAAC TTGGACAAAA TTAAACAAAT AGAAGGTATT 480

TCATCGGTTG AAAGGGCACA AAAAGTCCAA CCCATGATGA ATCATGCCAG AAAGGAAATT 540

GGAGTTGAGG AAGCTATTGA TTACCTAAAG TCTATCAATG CTCCGTTTGG GAAAAATTTT 600

GATGGTAGAG GTATGGTCAT TTCAAATATC GATACTGGAA CAGATTATAG ACATAAGGCT 660

ATGAGAATCG ATGATGATGC CAAAGCCTCA ATGAGATTTA AAAAAGAAGA CTTAAAAGGC 720

ACTGATAAAA ATTATTGGTT GAGTGATAAA ATCCCTCATG CGTTCAATTA TTATAATGGT 780

GGCAAAATCA CTGTAGAAAA ATATGATGAT GGAAGGGATT ATTTTGACCC ACATGGGATG 840

CATATTGCAG GGATTCTTGC TGGAAATGAT ACTGAACAAG ACATCAAAAA CTTTAACGGC 900

ATAGATGGAA TTGCACCTAA TGCACAAATT TTCTCTTACA AAATGTATTC TGACGCAGGA 960

TCTGGGTTTG CGGGTGATGA AACAATGTTT CATGCTATTG AAGATTCTAT CAAACACAAC 1020

GTTGATGTTG TTTCGGTATC ATCTGGTTTT ACAGGAACAG GTCTTGTAGG TGAGAAATAT 1080

TGGCAAGCTA TTCGGGCATT AAGAAAAGCA GGCATTCCAA TGGTTGTCGC TACGGGTAAC 1140

TATGCGACTT CTGCTTCAAG TTCTTCATGG GATTTAGTAG CAAATAATCA TCTGAAAATG 1200

ACCGACACTG GAAATGTAAC ACGAACTGCA GCACATGAAG ATGCGATAGC GGTCGCTTCT 1260

GCTAAAAATC AAACAGTTGA GTTTGATAAA GTTAACATAG GTGGAGAAAG TTTTAAATAC 1320

AGAAATATAG GGGCCTTTTT CGATAAGAGT AAAATCACAA CAAATGAAGA TGGAACAAAA 1380

GCTCCTAGTA AATTAAAATT TGTATATATA GGCAAGGGGC AAGACCAAGA TTTGATAGGT 1440

TTGGATCTTA GGGGCAAAAT TGCAGTAATG GATAGAATTT ATACAAAGGA TTTAAAAAAT 1500

GCTTTTAAAA AAGCTATGGA TAAGGGTGCA CGCGCCATTA TGGTTGTAAA TACTGTAAAT 1560

TACTACAATA GAGATAATTG GACAGAGCTT CCAGCTATGG GATATGAAGC GGATGAAGGT 1620

ACTAAAAGTC AAGTGTTTTC AATTTCAGGA GATGATGGTG TAAAGCTATG GAACATGATT 1680

AATCCTGATA AAAAAACTGA AGTCAAAAGA AATAATAAAG AAGATTTTAA AGATAAATTG 1740

GAGCAATACT ATCCAATTGA TATGGAAAGT TTTAATTCCA ACAAACCGAA TGTAGGTGAC 1800

GAAAAAGAGA TTGACTTTAA GTTTGCACCT GACACAGACA AAGAACTCTA TAAAGAAGAT 1860

ATCATCGTTC CAGCAGGATC TACATCTTGG GGGCCAAGAA TAGATTTACT TTTAAAACCC 1920

GATGTTTCAG CACCTGGTAA AAATATTAAA TCCACGCTTA ATGTTATTAA TGGCAAATCA 1980

ACTTATGGCT ATATGTCAGG AACTAGTATG GCGACTCCAA TCGTGGCAGC TTCTACTGTT 2040

TTGATTAGAC CGAAATTAAA GGAAATGCTT GAAAGACCTG TATTGAAAAA TCTTAAGGGA 2100

GATGACAAAA TAGATCTTAC AAGTCTTACA AAAATTGCCC TACAAAATAC TGCGCGACCT 2160

ATGATGGATG CAACTTCTTG GAAAGAAAAA AGTCAATACT TTGCATCACC TAGACAACAG 2220

GGAGCAGGCC TAATTAATGT GGCCAATGCT TTGAGAAATG AAGTTGTAGC AACTTTCAAA 2280

AACACTGATT CTAAAGGTTT GGTAAACTCA TATGGTTCCA TTTCTCTTAA AGAAATAAAA 2340

GGTGATAAAA AATACTTTAC AATCAAGCTT CACAATACAT CAAACAGACC TTTGACTTTT 2400

AAAGTTTCAG CATCAGCGAT AACTACAGAT TCTCTAACTG ACAGATTAAA ACTTGATGAA 2460

ACATATAAAG ATGAAAAATC TCCAGATGGT AAGCAAATTG TTCCAGAAAT TCACCCAGAA 2520

AAAGTCAAAG GAGCAAATAT CACATTTGAG CATGATACTT TCACTATAGG CGCAAATTCT 2580

AGCTTTGATT TGAATGCGGT TATAAATGTT GGAGAGGCCA AAAACAAAAA TAAATTTGTA 2640

GAATCATTTA TTCATTTTGA GTCAGTGGAA GCGATGGAAG CTCTAAACTC CAGCGGGAAG 2700

AAAATAAACT TCCAACCTTC TTTGTCGATG CCTCTAATGG GATTTGCTGG GAATTGGAAC 2760

CACGAACCAA TCCTTGATAA ATGGGCTTGG GAAGAAGGGT CAAGATCAAA AACACTGGGA 2820

GGTTATGATG ATGATGGTAA ACCGAAAATT CCAGGAACCT TAAATAAGGG AATTGGTGGA 2880

GAACATGGTA TAGATAAATT TAATCCAGCA GGAGTTATAC AAAATAGAAA AGATAAAAAT 2940

ACAACATCCC TGGATCAAAA TCCAGAATTA TTTGCTTTCA ATAACGAAGG GATCAACGCT 3000

CCATCATCAA GTGGTTCTAA GATTGCTAAC ATTTATCCTT TAGATTCAAA TGGAAATCCT 3060

CAAGATGCTC AACTTGAAAG AGGATTAACA CCTTCTCCAC TTGTATTAAG AAGTGCAGAA 3120

GAAGGATTGA TT 3132

(2) INFORMATION FOR SEQ ID NO: 111:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14672 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

CGAGATTTCT TTAAATGAAC TACGTGAAAT CTACCCATCA TCCAGATCTG GATATTCTCT 60

CCTATCTATA AGTAAAGTTT TAGGAGATTT TAATATAAGT TCTCATGCTT TTAAAGCTTC 120

GGTAAGAGAT TTAAAACCGC TCAGTTTCCC ACTCATTTGC TTCTGGGAGA GTTCTCATTT 180

TATTATTCTT GAAAAAATTA GTAAAAACAA GTTTTATATT TTAGATCCTG CAAAAGGCAG 240

GCAGAGAATG TCAATAAGTG AATTTGAAAG GCATTATTCA AATATCATTT TAACATTTAA 300

AAAGTTAGAT AGCTTTATGT CTCGTAAAGA TAATAAGAAG TCGCCTGTTT TAAAGTATTT 360

TTTTAAGTAT AGGAATAAGC TAGGGATTTT ATTTTTTGTA ACAGCATTAT TGTATGTAAT 420

ACAATCATTA GTACCTATAG CTAATAGATA CATAATTGAC ACGAATTTCA AGGACGATTC 480

GTATTCGTCT AGAATGTTAT TTACTATATT ATTTATATTT ACTGTTTCAT TCTCACTAAT 540

GTATTTATTA AGACAGATAT ATGTTGCATC CTTAAAATAT ATAATGGATA AAGAGATTAG 600

CTATGATTTT ATGAAACATT TGATATATTT ACCTTACAGT TTTTATGAAA AACGTACTTT 660

AGGGGATATA CTTTTTAGAG CTAACTCTAT TGTTTATATA AGAGAAATAC TATCAAATAA 720

TTTTATAGCA GCTATACTTG ATTTGTTAAT GATTGTGGTT TATGCTGTGG TTTTATTTAG 780

CTTTTCTAAG TACATGGTAA TCTTTTTAAT ATCACTAAGT CTAGCTCTAT CTATTGTAAT 840

GTATCCAATC ATAAAAATCT CAAAAAATTT AATTGATAAA AATATAAAAG AAAAGGTTAA 900

TGTTCAAAAT ATTACTTCCG AAGTAATTTC TAAAAATAGT GATATTAAGC TAACTGGAGA 960

AGAGGAATTT TGGATTAACA AATGGGATAA TTTTAATACA AAACAGCTCA TCATAGGTCG 1020

AAAACTTGAT ATACATTTAT CAATTGTTAG TAGTATAACG AATGTTTTAC AAATTATTCT 1080

CCCTGTTTTG ACCCTTATTG TAGGTGTAAA TATAAAAACA TTCGAACAAT TGACGTTAAG 1140

ACAAATTGTA GCAATAAGTA CAGTCTCACC ATACTTTATT TCTCCTATAA TTTCTTTAAG 1200

TGATAACTAT ATACAATTAA TGTTATTAAA GGGATATTTT TTAAGAATAG AGGATGTGTT 1260

TAATACTAAA TCCGAATTAA TTCCAGAAAG AGTCAGTCAA GATATAAAAT TTGATAAAAA 1320

AATAGAATTA AAAGATATTT GGTATAAATA TGGATTATTT GATGATTATG TTTTGAAAGG 1380

AATAAATGTT ACTATTAAAA AAGGAGAAAC TGTTGCTATT GTTGGAGAAT CAGGTTCAGG 1440

TAAGAGTACA TTAGCTAAAA TTTTATTAGG TTTATTAGAA CCTAATATTG GTTCAATAGA 1500

AGTTGATGGA GTAGAAAAAG AAGAAATTGG TCAAACATTG TATAGAAAGA TTTTTGGAGC 1560

AGTGTTACAA AATTCAACCC TAAGTTATGG TACCTTAAGA GAGAATTTGA CATTTGGACA 1620

CTTTGTTTCA GATGAAGAAT TAATGACAAA TCTAAATTCA ATTGGTCTTA GCAATGTAGT 1680

TAAATCTTTA CCTCTTGGAT TAGAGACAAT CATCGCTGAA GAAGGTAATA ACTTTTCTGG 1740

AGGGCAGCAG CAAATGATAC TTTTAGCTCG TTGTCTTTTG TCGAAACCTT CGGTAGTTGT 1800

TTTGGACGAA GCAACAAGTA GTTTAGATAA TTTATCTCAA CAAATTACAA CTTCTTACTT 1860

AAGTGAAATC GGTACCACTA AGATTTTAAT TGCCCATCGA CTAGATACTA TCAAGTCTGC 1920

AGATAAGATC TTAGTAATGC ATAATGGTGA AATTGTAGAG ATTGGGACCC ATAGAGAACT 1980

TCTTGAACTA GGAGGCATTT ATAAGCAATT GTATTCAAAT AATTAGTTTT TGATTAAAAG 2040

GGTAAATTTA TGAAGATTAT GAAAAAAAAA TATTGGACTT TAGCGATATT ATTCTTTTGT 2100

TTGTTCAATA ATTCTGTTAC TGCTCAAGAA ATACCTAAAA ATCTTGATGG CAATATAACT 2160

CACACTCAGA CTAGCGAAAG TTTTTCTGAA TCTGATGAAA AACAGGTTGA CTATTCTAAT 2220

AAAAATCAAG AAGAAGTAGA CCAAAATAAA TTTCGTATTC AAATCGATAA GACAGAATTA 2280

TTTGTAACAA CAGATAAACA TTTAGAAAAA AACTGTTGTA AATTGGAACT TGAACCACAA 2340

ATAAATAACG ATATTGTTAA CTCTGAAAGT AATAATTTAC TAGGCGAAGA TAATTTAGAT 2400

AATAAAATTA AGGAAAATGT TTCTCATCTA GATAATAGAG GAGGAAATAT AGAGCATGAC 2460

AAAGATAACT TAGAATCGTC GATTGTAAGA AAATATGAAT GGGATATAGA TAAAGTTACT 2520

GGTGGAGGCG AAAGTTATAA ATTATATTCT AAAAGTAATT CTAAAGTTTC AATTGCTATT 2580

TTAGATTCAG GAGTCGATTT ACAAAATACT GGATTACTGA AAAATCTTTC AAATCACTCA 2640

AAAAACTATG TCCCCAATAA AGGATATTTA GGAAAAGAGG AGGGAGAGGA AGGAATAATA 2700

TCAGATATTC AAGATAGATT AGGTCATGGT ACGGCTGTTG TAGCTCAAAT TGTAGGGGAT 2760

GACAATATTA ATGGAGTAAA TCCTCACGTT AATATTAACG TCTATAGAAT ATTTGGTAAG 2820

TCGTCAGCTA GTCCAGATTG GATTGTAAAA GCAATTTTTG ATGCTGTAGA TGATGGCAAT 2880

GATATTATCA ATCTTAGTAC TGGACAATAT TTAATGATTG ATGGAGAATA TGAGGACGGA 2940

ACAAATGATT TTGAAACATT TTTGAAGTAT AAAAAGGCTA TTGATTACGC GAATCAAAAA 3000

GGAGTAATTA TAGTAGCTGC ATTAGGGAAT GACTCCCTAA ATGTATCAAA TCAGTCAGAT 3060

TTATTGAAAC TTATTAGTTC ACGCAAAAAA GTAAGAAAAC CAGGATTAGT AGTTGATGTT 3120

CCAAGTTATT TCTCATCTAC AATTTCGGTC GGAGGCATAG ATCGCTTAGG TAATTTATCA 3180

GATTTTAGCA ATAAAGGGGA TTCTGATGCA ATATATGCGC CTGCAGGCTC AACATTATCT 3240

CTTTCAGAAT TAGGACTTAA TAACTTTATT AATGCAGAAA AATATAAAGA AGATTGGATT 3300

TTTTCGGCAA CACTAGGAGG ATATACGTAT CTTTATGGAA ACTCATTTGC TGCTCCTAAA 3360

GTTTCTGGTG CGATTGCAAT GATTATTGAT AAATACAAAT TAAAAGATCA GCCCTATAAT 3420

TATATGTTTG TAAAAAAATT CTGGAAGAAA CATTACCAGT AAAAAATGGT ATAAAAGTGT 3480

TAAATATACC AAACGTATTG AGATATGATT TGAATATGTT ACAATTAGAA TATAAAAATG 3540

AACAAAGTTG GGATAGTTTC ATAGATAATG TTAATTTAAT TGAGTTGGAA GAGAGAATTC 3600

AAACTACTAT TGGAATTAAA CAAATAAACA CACACAATAT TATTACTATT GCCCGAGAAG 3660

GGTACTCTCA AAATTATTTA CCTAACACTT CAGAAAATAC ATATAATTCA TTACAAGTCA 3720

GTTTAGTTGG AGTATTACTA CTTTTTATAA GTATGGTAAA TATTTTATGG GCTAAAAAAA 3780

GTAAATGAAA ATAAAATTTG GAGCCCTCTG AAAAAGTAAG TCCTACAGTT CAACTAAAAT 3840

GAGTCAAAAG ATGAATCACC TTGATGTAGG GGAGTTTGTC TTATTGCTGC CTGAACACCT 3900

CCGTTCAGAG GAAGAACATT ATAAATCTGT TTTTGAAGAC GACTTAACCA GTCGCATATC 3960

TAGTCAAGAT GAACGACAGC AAATGACTGC TACGGTAGGT TATTTAGAAT CAGGTCAGGA 4020

TCGTTTTGTG TATAATACGA CCCCTATTTC TTACCAGCAG TTTTTGAAAG ATCCAATCAT 4080

CATTGTTATA ACACCCCAAT CAACTGGTCC ACAGTCCATT TTGTTTTGGA TAGACGCAGT 4140

ACAGAACTAC GTTCTCTTTA ATCAATTGTC TGATGCCCAG GAGCTTATCC AGAGACAAGG 4200

CATTGAAAAT TGGGTCTCAG AAATGCAAAC AGGTTACCAC AACTACATCA CATTATTGGA 4260

TAATATCCAG AGGGAACGTT GGGTAATGCT AGCAGGAGCT GTGCTTGGGA TTGCAACTTC 4320

AATCTTGTTG TTTAACACTA TGAATAGGCT CTACTTTGAA GAATTTAGAC GTGCCATTTT 4380

TATCAAACGC ATTGCAGGTC TCAGGTTCTT AGAAATCCAT CGCACTTATC TCTTTGCTCA 4440

ACTGGGTGTG TTTTTACTGG GATTTGTTGC GAGTGTATTT CTTCAGGTAG AGATAGGAGT 4500

TGCTTTCTTA GTCTTGTTAC TCTTTACTGG TCTATCTCTT TTACAGTTAC ATGTCCAAAT 4560

GCAGAAAGAA AACAAGATGT CCATGCTTGT TTTGAAGGGA GGTTAATATG ATTGAACTTA 4620

AACAGGTGAG TAAATCTTTT GGAGAACGAG AGTTATTTTC GAATCTTTCA ATGACATTTG 4680

AGGCTGGAAA AGTCTATGCC TTAATTGGTT CAAGTGGTAG CGGAAAAACA ACCTTGATGA 4740

ACATGATTGG GAAATTAGAA CCTTATGATG GGACGATTTT TTACCGAGGT AAAGACTTGG 4800

CCAATTATAA ATCAAGTGAT TTTTTCCGTC ACGAATTGGG CTACCTCTTC CAGAACTTTG 4860

GCTTAATTGA AAACCAAAGT ATTGAAGAAA ACCTTAAGCT AGGTCTCATT GGTCAAAAGT 4920

TGAGTCGGTC GGAACAGCGG TTGAGGCAGA AGCAGGCTTT AGAACAGGTC GGCCTGGTTT 4980

ATCTTGACCT AGATAAGCGC ATCTTTGAGT TATCGGGCGG AGAATCGCAA CGGGTTGCCT 5040

TGGCAAAAAT TATCTTAAAG AATCCACCCT TTATTCTGGC AGATGAGCCA ACAGCTTCAA 5100

TAGACCCAGC AACCTCTCAG TTGATTATGG AGATTTTGCT ATCTCTTCGA GATGATAATA 5160

GGCTAATCAT TATCGCAACA CATAATCCGG CAATTTGGGA GATGGCTGAT GAAGTGTTCA 5220

CGATGGATCA TCTGAAATAA AAATCCTTGT TTTTAATTGC ACGATGAGTT ACTGAAATAT 5280

TATCATGAAT CAAGAATTGG AGTTAATTTA GAATTGTACT TAATTTAGAA TTGTACTTTA 5340

TTAATATTGA GGTAACTTTT TCTTGATAAA GGAAGAAATA ATGGAGAGGA AGTTAGAATG 5400

AAAAAATTCG ACAATTATAT TATTGAGAAG CCTTGCGATT CTAATTCAGA TAAACTGCAA 5460

AAAATCTTAA TAATTGAAAG TTTGGTAGAT GATATTTTGC AATTTTCTCT CAGAATCAAT 5520

AATAGTGTAG GAGAGATTTT CCTCCTACAA CCGTTTTAAA AGAAAACTAT CTTTATTCCA 5580

TGTTATTTTG AGGAAGATAT TGTGAAAGTC AAAGATGATG ATAAAGTTGA GTGGAATTTG 5640

TTAGAATTTC AAAAATTTAG AGCATTTTTG GCTTAGTAAT CTGTGTTGAA GGCTCAAAAC 5700

CTATGGTAAA AAAGTAGCTT TGAAAACGTA TTGCCTCCAA AGATTTAGTT AAATAATGAT 5760

TTAACACAAA AAGAAATTAT TGAAGTTCTG GAAAGATGTT GTTTCAGTAT TGAGAAAAGG 5820

TGGGAAAAAC TTGCGATTTT CACAGAGAAA GGAAGAAAAA GTATAGAAAT ATAGTCAATT 5880

GAAACAAGAA CAGGATAAAA GAACCTTTTG TGCCATATTT TTCTCCTTTC GCTTTACAAT 5940

TGGATTGAAC ACCTTTATTG TATCGCGTTT GGAGTTTTTT TGGTATAACC TTCGACGCAC 6000

ACCCGCATAG CGGGTGTTTT TTTTGTCTCG CACCTAACGG AGCGAGACAA ACTAATAGTC 6060

ACTTAATCAA AAAACGCACC ATATCAAAAA CTAAAAAGTT TGATATCATG CGTCATGTCT 6120

TAAACTAATT GACTATACTT TCTATTCAAA TGAGCTTTTA ACCAATTGAT TGAGCCAATC 6180

CACTCTTAAA ACCAAAGAGC AATTTCTCGC TTAGCTGACT CTTCTGAATC TGAACCATGT 6240

ACAACATTTT GGATAATCTC ATTTTCTCCA GCAGCTTTTG CAAAATCACC TCGAATAGTG 6300

CCTGGTAAAG CTTCTTCTGG ACGAGTTGCA CCCATCATGG TCCGCCAAGT TTCGATTACT 6360

TTGGGACCAG AAATGACACC CACAAGAACT GGACCTGAAG TCATGAATTC ACGAATCGGT 6420

GGGTAAAAAC TCTGACCAAC CAAGTCCTGA TAGTGCTGGT CAATCAACTC TTCTGAAACC 6480

TGTGAACGAA ACTCCAATTT TTCGATTGTA AATCCACGTT GTTCGATGCG CTTTAACACT 6540

TCACCCACTA GCCCTCTTTT TACACCATCT GGTTTGATGA TAAAGAATGT TTGTTCCATA 6600

CCCGTCTCCT TTGTCAGCTT CTTTCTTTTA TTTTACCACA TTTCGTGGAA AAATGGAGAA 6660

AGTTTTCAGA AGAGAGAATG AGAGAACCCT CGGGTTCTCT CATTCTCTCT TATTCTACTG 6720

TTTCTTCCAC AGTTTCAACG GCAGTATCCA CAACTACTTC TGTTGTTTCT TCATTTCCTT 6780

CTTCCTCTAC TGGAGGATTA AGGTATTCTT CTTCGTTGAC AGCATGTGGT TCAAGGTTAC 6840

GGTAACGGGC CATACCAGTA CCAGCTGGGA TGATCTTACC GATGATAACA TTTTCTTTAA 6900

GTCCAAGGAG ATGGTCTTTC TTACCACGGA TAGCTGCGTC AGTAAGGACA CGAGTTGTTT 6960

CCTGGAAGGA AGCCGCTGAC AAGAAACTGT TTGTTTCAAG TGAGGCTTTG GTAATTCCCA 7020

TAAGGACTGG GCGACCTGTC GCTGGAACTC CACCTGCGAT AAGGACATCT TTGTTGGCAT 7080

CTGTAAAGTC ATTGATATCC ATGAGGGTAC CCATGAGAAG ATCTGTATCA CCTGGATCCA 7140

TGACACGGAC TTTACGGATC ATTTGACGAA CCATTACCTC GATGTGTTTG TCACCGATTT 7200

CTACCCCTTG GCTACGGTAA ACTTTTTGTA CTTCACCGAG AAGGTACGTT TCAACTGACA 7260

AGACATCACG AACTGCAAGG AGACGTTTTG GTTGGATAGA ACCTTCTGTC AGAGCAGCAC 7320

CACGCGCTAC TTGGCCCCCA ACTTCGACAC GCATACGAGC TGTAAATGGA ACGACATATT 7380

CACCTTCGCC AGTTTCACCC TTAACAAAGA CTTTCTTGGT ACGAGTTGAT GCATCTTCTT 7440

CGATAGCAGT AACTTGTCCT TTAACCTCTG TAATAACCGC TTCCCCTTTA GGATTGCGGG 7500

CTTCAAAGAT TTCTTGGACA CGAGGAAGAC CCTGAGTGAT ATCGGTATTT GAGGCAACCC 7560

CACCTGTGTG GAAGGTACGC ATTGTAAGCT GTGTACCAGG TTCCCCGATA GATTGGGCAG 7620

CGATTGTACC AACTGCTTCA CCAACTTCAA CCGCATCACC AGTCGCCAAG TTGATACCGT 7680

AACAGTGACG GCAGACACCG TGACGAGTGT TACATGTAAA TACAGAACGG ATAGTCACTT 7740

CTTCCACACC AGCATTGACA ATTTCACGCG CCTTGTCTTC TGTAATCAAT TCATTTGGAC 7800

CAATAATCAC TGCACCAGTT TCTGGATGTT TAACAGTTTT CTTAGTGTAA CGACCGTTGA 7860

GACGCTCTTC GAGAGACTCG ATCATCTCTT TTCCTTCTGC GATAGAACGG ATCAAGAGAC 7920

CACGGTCAGT TCCACAGTCG TCCTCACGGA TGATAACGTC TTGGGCAACG TCGACCAAAC 7980

GACGAGTCAA GTAACCTGAG TCGGCTGTCT TAAGGGCCGT ATCGGTCATA CCTTTACGAG 8040

CACCGTGAGT TGAGAAGAAC ATTTCCAATA CCGACAAACC TTCGCGGAAG TTTGAAAGGA 8100

TTGGCAATTC CATGATACGT CCATTCGGAG CAGCCATCAG ACCACGCATA CCGGCAAGCT 8160

GTGAGAAGTT TGAGATGTTA CCACGGGCTC CAGAGTCCAT CATCATAACG ATTGGGTTCT 8220

TAGGATCTTG GTTAGCAATC AAGCGTTTCT CAAGTTTTTC ACGGGCAGCA CGCCATTCAG 8280

CTGTAACAGC ATTGTAACGC TCGTCGTCTG TGATCATACC ACGACGGAAT TGTTTGGTGA 8340

TTTGTTCGAC ACGTTTGTGT GATTCTTCAA TGATTTCAGC CTTGTCATCA ACGACTGGGA 8400

TATCGGCAAT ACCCACTGTC AATCCTGCAA GAGTTGAGTG GTGGTAACCG AGGTTCTTCA 8460

TGCGGTCAAG TAGGGCAGAA GTTTCTGTCG TACGGAAACG TTTGAAGATT TCAGCGATGA 8520

TATTTCCAAG GTTTTTCTTC TTGAATGGAG GGTTGAGCTC AAGATTGCTG ATAGCTTCCT 8580

TGATATCTCC ACCAAGTGGC AAGAAGTATT TAGCTGGAAC ACCTTCTGTC AAGTTGGCAT 8640

TGTTTGGTTC TTGCAAGTAT GGTAGCCCCT CTGGCATGAT ATCGTTGAAG AGAATTTTAC 8700

CAACTGTTGT AAGCAAGACC TTATGTCTTT GCTCTTCTGT CCAAGGCTTG TTGAGGCTGT 8760

CTGTTGCGAT ACCAACACGT GAGTGGAGGT GAACATAACC ATTGCGGTAA GCCATAACCG 8820

CTTCGTCACG GTCTTTGAAG ACCATTCCTT CACCTTCGCG ACCAGCTTCT TCCATGGTCA 8880

AGTAGTAGTT ACCCAAAACC ATGTCCTGAG ATGGAGTAAC TACCGGTTTC CCATCTTTCG 8940

GGTTCAAGAT GTGCTCAGCA GCTAGCATGA GGATACGAGC TTCTGCTTGT GCTTCTTCTG 9000

AAAGTGGTAC GTGGATGGCC ATTTGGTCCC CGTCAAAGTC AGCATTGTAG GCTTCACAGA 9060

CAAGTGGGTG CAAGCGAAGA GCCTTACCAT CAATCAAGAC TGGCTCGAAG GCTTGGATAC 9120

CCAAACGGTG AAGGGTCGGT GCGCGGTTCA AAAGCACTGG GTGTTCTTTA ATCACTTCTT 9180

CAAGGATATC CCAGATACGC TCATCTCCGC GTTCCACCAA GCGTTTAGCT GCTTTGACGT 9240

TTTGCACGAT ATCACGGGCA ACGATTTCAC GCATGACAAA TGGTTTAAAG AGTTCAATCG 9300

CCATTTCACG CGGCACACCA CATTGGTACA TCTTAAGAGT TGGACCAACG GCGATAACTG 9360

AACGTCCTGA GAAGTCAACA CGTTTACCGA GCAAGTTTTG ACGGAAGCGT CCTTGTTTAC 9420

CTTTAAGCAT GTGGCTCAAT GATTTCAATG GACGGCTACC TGGTCCTGTG ATTGGACGAC 9480

CACGACGACC ATTGTCAATC AAAGCGTCAA CTGCTTCTTG AAGCATACGC TTCTCATTTT 9540

GAACGATGAT ACCTGGTGCA TTTAACTCAA GCAAACGAGC CAAACGGTTG TTACGGTTGA 9600

TAACACGGCG GTAAAGGTCA TTCAAGTCAG ATGAGGCAAA ACGGCCACCA TCCAACTGCA 9660

ACATTGGACG AAGATCTGGT GGGATAACCG GAAGGATGTT AAGAATCATC CATTCAGGTT 9720

TGTTTCCAGA CTTGTAAAAG GCATCCAAAA CATCCAAACG ACGGATGGCT TTGACACGCT 9780

TTTGTCCAGT AGCTGTTTTC AATTCTTCTT TGAGTTCAGC AATTTCTTTT TCAAGATCTA 9840

CTTGCTTCAA AAGGTCTTGG ATGGCTTCCG CACCCATCTT GGCAACAAAT GAACCATAAC 9900

CATATTCACG CAAGCGCTCT CGGTATTCGC GCTCTGTCAT GATAGACTTG TGCTCAAGTG 9960

GTGTATCCTT AGGATCAATC ACCACATAAG CCGCAAAGTA GATAACTTCC TCGAGGGCAC 10020

GAGGGCTCAT ATCAAGGGTC AAGCCCATAC GGCTTGGAAT CCCCTTGAAG TACCAGATGT 10080

GAGATACAGG AGCTTTCAAT TCGATATGTC CCATACGCTC ACGACGAACT TTCGTACGCG 10140

TTACTTCAAC CCCACAGCGG TCACAAACAA TTCCTCTGTA ACGAATGCGT TTGTACTTAC 10200

CACAAGCACA TTCCCAGTCT TTTGTAGGAC CAAAGATCAC TTCATCAAAG AGTCCTTCAC 10260

GTTCTGGTTT CAAGGTACGA TAATTGATTG TTTCAGGTTT TTTGACTTCT CCATAAGACC 10320

ATGAACGGAC TTTACTTGGA GAAGCTAGGG TGATTTGCAT ACTTTTAAAA CGATTTACAT 10380

CAACCACTAT TTCTTCCCTT TCTATTCTAA GTGAACTGCT TATTCTTGTT CAGCAGCTTC 10440

TTCTGTTGCT TCCGCTTTTG TTGCTTTCTC AGCTTCTTCA GCTTCAAAGG CTGCTTTAGC 10500

CTCTTGGGCT GCTTTTTCGC GGGCTTTTTC AAGGTCATCT ACGTGGATGA CATCTTCGTC 10560

CATTCCTTCA TCCAAGTCGC GAAGTTCCAC TTCTTGGTCA TCTTCGTCTA GGACACGCAT 10620

GTCAAGACCA AGAGATTGCA ATTCTTTGAC AAGAACTCGG AAGGATTCTG GAACACCTGG 10680

TTTTGGAATT GGTTTGCCTT TTGTAATAGC TTCATAGGCT TTCAAACGTC CGTTGATATC 10740

GTCCGACTTG TAAGTCAAGA TTTCTTGAAG GACATTTGAC GCACCGTAGG CTTCAAGAGC 10800

CCAAACCTCC ATCTCACCGA AACGTTGTCC ACCAAACTGA GCCTTACCTC CGAGTGGTTG 10860

TTGGGTAACA GTTGAGTATG GTCCGACTGA ACGCGCGTGC AATTTATCAT CAACCATGTG 10920

GTGGAGTTTG ATCATGTACA TGACTCCGAC AGAAACACGG TTATCAAACG GTTCACCAGT 10980

ACGTCCATCG TAAAGGATCG TTTTGGCATC GCTATCCATA CCTGCTTCTT TAACAGTTGA 11040

CCAAAGATCT TCAGAACTTG CTCCATCAAA GACTGGTGTA GCGATGTGAA TACCAAGAGT 11100

ACGAGCTGCC ATACCAAGGT GAAGCTCCAT AACCTGACCG ATATTCATAC GTGATGGTAC 11160

CCCAAGTGGG TTCAACATGA TGTCGACTGG AGTTCCGTCT GGAAGGTAAG GCATGTCTTC 11220

TACAGGAACG ATACGAGAGA CAACCCCTTT GTTTCCGTGA CGTCCGGCCA TTTTATCTCC 11280

GACCTTAATC TTACGTTTTT GAGCGATGTA AACACGAACC AACATGTTAA CACCTGATTG 11340

CAACTCATCT CCATTTACAC GTGTAAAGAT CTTAACATCA CGAACGACAC CATCGGCACC 11400

GTGTGGTACA CGAAGAGAAG TATCACGCAC TTCACGAGAC TTGTCTCCAA AGATAGCGTG 11460

CAAGAGACGT TCTTCAGCTG AAAGATCTTT CTCACCCTTA GGTGTTACTT TACCTACAAG 11520

AATATCACCT TCTTTAACCT CAGCACCAAT ACGGATAATC CCCATTTCGT CAAGGTCTTT 11580

GAGGGCATCT TCACCAACGT TTGGAATTTC GCGAGTGATT TCTTCAGGCC CAAGCTTTGT 11640

ATCGCGCGTT TCTGATTCGT ATTCTTCAAG GTGAACAGAT GTGTAGACAT CGTCCTTCAC 11700

CAAGCGTTCG CTCATGATAA CGGCATCCTC GAAGTTGTAA CCTTCCCAAG TCATGTAGGC 11760

AACGATTGGG TTTTGTCCAA GCGCCATTTC TCCATTTTCC ATAGAAGGTC CGTCAGCGAT 11820

GAAATCGCCT TTTTCAACGA CATCACCAAC TTTTACGAGA GTGCGTTGGT TGTAAGCAGT 11880

ACCTGAGTTT GAACGACGGA ATTTTTGGAT GTGGTAAACA TCCAATGAAC CATCTTCACG 11940

ACGAACTTCT ACCTTGTCAG CATCTGCGTA AGTAACTTTA CCATCATACT GAGCAATCAC 12000

AGCCGCACCA GAATCGTGGG CTGCTTGGTA TTCCATACCA GTACCAACGT AAGGTGCCTG 12060

AGGATTAATC AATGGCACAG CCTGACGTTG CATATTGGCT CCCATGAGGG CACGGTTGGA 12120

GTCATCGTTT TCCAAGAAAG GAATACATGC TGTCGCAACG GCAACTACCT GTTTTGGTGA 12180

AACGTCCATG TAGTCAACAA TATTAGCTGG ATACTCTTGG TTGACCCCTT GGTGACGTCC 12240

CATGACAATC TTCTCAGCAA AGGTTCCATC TTCATTCAGA CGAGAGTTAG CCTGAGCTAC 12300

AGTATATTCA TCTTCTTCAT CAGCTGTCAA CCAAACAATT TCGTTCGTGA CAACACCTGT 12360

TTCACGGTCA ACCTTACGGT ATGGTGTTTG AACAAAACCA TATTTGTTCA AGTGTCCATA 12420

AGATGACAAG TTATTGATCA AACCGATGTT AGGTCCTTCA GGTGTCTCGA TTGGACACAT 12480

ACGACCATAG TGAGTGTAGT GCACGTCACG TACTTCATAT CCAGCACGGT CACGAGTCAA 12540

ACCACCAGGT CCTAAGGCTG ACAAACGGCG TTTGTGAGAC AACTCAGAAA GCGGGTTGTG 12600

TTGGTCCATG AACTGTGACA ACTGTGATGA ACCAAAGAAT TCTTTAACTG CAGCTGTTAC 12660

AGGACGGATA TTGATAATTT GTTGTGGTGT CAAGACTTCA TTGTCCTGAA CAGACATACG 12720

TTCACGGACA TTACGTTCCA TACGAGAAAG TCCCAAACGT ACTTGGTTGG CAAGCAATTC 12780

ACCAACCGCA CGGATACGAC GATTTCCAAG GTGGTCGATA TCATCTACAC GGCCAAGTCC 12840

TTCAGCCAAG TTGAGGAAGT AGCTCATCTC AGCAAGGATA TCTGCAGGAG TCACCGTACG 12900

AACCTTGTCA TCTGGGTTAG CATTACCAAT GATCGTTACG ACGCGATCTG GATCAGTTGG 12960

AGCAATAACC TTGAATTTTT GAAGAACAAC AGGCTCAGTC ACAACGGCTG CATCGTTTGG 13020

GATGTAGACA ATCTTGTTCA AGTCGCCATC CAAATGGCTT TCAATGCTTT CAATCACGCT 13080

ACGAGTCATA ATCGTACCAG CTTCTACCAA GATTTCTCCA GTTTCAGGGT CTACCAATGG 13140

CTCTGCAATG GTTTGGTTGA GCAAACGTGT TTTAACATTG AGTTTTTTAT TGATTTTGTA 13200

ACGACCAACT GCTGCCAAGT CATAACGACG TGGGTCAAAG AAGCGAGCTA CAAGCAAGCT 13260

ACGTGAGCTT TCAGCCGTCT TAGGCTCACC TGGACGAAGG CGTTCGTAAA TTTCTTTCAA 13320

GGCTTCGTCT GTACGAGAGT CCATTGGATT CTTGTGGATA TCTTTTTCAA CAGTGTTGCG 13380

AACCAATTCG CTGTCACCAA AGATATCAAA GATTTCATCA TCACCTGAGA AACCAAGAGC 13440

ACGAACCAAG GTTGTAAATG GAATCTTACG AGTACGGTCG ATACGAGTGT AGGTGATATC 13500

TTTTGAGTCG CTTTCAAGTT CCAACCAAGC TCCACGGTTA GGGATAACAG TTGAACCATA 13560

GCCCACCTTA CCATTTTTGT CTACTTTGTC GTTAAAGTAA ACACCTGGTG AGCGGACCAA 13620

CTGAGAAACG ATAATACGTT CACCACCATT GATGATGAAA GTACCCATTT CTGTCATGAT 13680

TGGGAAATCA CCAAAGAAAA CTTCTTGGGT CTTGATTTCG CTTGTTTCTT TATTGATCAA 13740

ACGGAAGGTT ACAAAAATTG GTGCTGAGTA GCTAGCATCG TGGATACGAG CTTCTTCTAG 13800

CGTATATTTT GGTTCCTTGA TTTCATATCC AACAAATTCC AACTCCATTG TGTCTGTGAA 13860

GTTTGAAATT GGCAATACAT CTTCAAACAC TTCCTTAAGA CCGTGGTCTA GGAAAGCTTT 13920

GAATGAGTCA GTTTGAATTT CAATCAAATT TGGTAAGTCA AGAACTTCTT TGATTCTTGA 13980

AAAACTACGA CGGGTACGAT GTTTCCCGTA TTGAACGTCA TGTCCTGCCA AGATGATTCT 14040

CCTTTGTAAA TAAGTTCCAA GCCTTGTCAA TCAGGCTTTT CTAATCGTCA TATGGTTGTA 14100

AACCCCTTAT CACCGTGTCC TCTTGACGAA TTTTCAGAAT CTTTAAGCCT CTGTTACAAA 14160

TGCTCAAAAT CTTGAAAAAA AGCACAAAAA GAGCAGCTAA ATCTGACTTT TTCAGAAGAT 14220

TTAACTGCTG TGAGCCTTGT CTGGACAATA TTTCAGACAA AACCTACGAC AAATGATTAC 14280

CCATATTATA CCCTATTTAG CTAGATTTTT CAAGGGGTTT CAGTAGGTTT TTGGTAAATT 14340

TTTTCCCATA GAAAACTTGG CATCACATTC GAATCACGCT ATGGTACAAA AAACTGAAAA 14400

AACTATTGAC TGAAAATCAT TTTCAAGGTA TAATAATAAA CGTTAAGGCG GTATAGCCAA 14460

GTGGTAAGGC ACGGCTCTGC AAAAGCTTGA TCGTCGGTTC AAATCCGTCT ACCGCCTTCT 14520

ATAACTTGAT TTATCAGGTT TCAAATGAAC AGAAAGCCCA ATTTGAAGGG CTTTTTTTAT 14580

TTTCCCTCGA ATAAATACGT ATAACTTTAA AAACTTTTGG AGCGAGTTTG TGGCAGAGTT 14640

CTTTCCATGG CATAATTCCC TTTTGAAATC AG 14672

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7902 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

AGGAGACTAT TCAAGCCCAA ATTgAGTAGC CCAGCAAAGA CTGTATAGAC TGTGATACGT 60

TTTTCATAGC CATTGGTAAA GAGAATTTGG GAACCAAGAA TGGTATCTAA GGCCAGGATA 120

ATCGTACGAA AAGCGAAGAG AGAGGTCAAG ATGCCGCCTC CGATATATTT TTCACTACCG 180

TAAAGTAGGA TGGCATTTGG TCCTAAAACC ATGAGTCCAA AACTCAGTGG AATGATAAAG 240

AAGTTAAAGA TTCGACTACC TCTATTAACC AGAGAAACAT AGGCTTCTTT GTCTCCTTTC 300

CCCAGATAGT AACTGAGACG AGGCACACTC ACTCCAATTG CACCTGTTAC AACCCCAGCT 360

ATAACGGTCA CAATTCGCTG AGCTATGGTA TAGTAACTAA CGTTGACATC AATCCCTGTT 420

TTAACGAGGA AGAGGCGATC TAAAAAAGTG AAGAGCATAT TGGCATTGGC AAAGACTAAC 480

ATGGCTGTCA GAGGGAGAAA GAGTGGTTTA AAATCACTTA GGTGAATTTT AACAAGTTTG 540

ATGTCTCTTT TAATCCAAAA ATAACTAATC AGGTAGTTAA TCAGCGTCGA TAAACTCATC 600

ACAAGTGTAT AGACAACAAT ATCGTGTTCA TTTTTAACAA ATAAGAAAAT AGAGACCAGC 660

ATCAGGATAC GGATGAAGGC AGTTTTGTAA AAGAGAAAAC TGTAATTTTC CAGAGCTTCA 720

TTGACCCATT CGATTGAAAA AATCTGGGCA ATGAGTTGAA TCCCCATAAC AAGGTAGACC 780

TTTTTGACGA TTGGATTATC AGTAAAGAAG AGAGGATAGG CTAGGATATA GACAGCAGTG 840

GTCAAAATCG TACAAGCGAT GCACAAATAA AAAAGACTAG AAAAGGTTCT GTTAAGATCT 900

TTTTTGTTAT CCTTGACATT ACTGATAGCC CTTAAACCGT AGTTATAGAC ACCATAAGTT 960

GCAAAGGGCA AGAAAAATGA CAAAATAGTG TCGACTGAGT TGAAGTAACC ATAGTCAGTT 1020

CGGTCCAAGA CACGCGCGAC ATAGGTTCCA GTTAGGATGG GAAAAATAAT ATTCAAGACA 1080

CGAATTCCCA TGTAAGATAG AGCATTTAAT TTTATACTTT TCATTCAATT TACCTCGTTT 1140

TTCATTATAT CATAAAGTTA GCTAATAAGA AATGAAGGGC AGTAAGTCAA GTAATCACTT 1200

TGAAGTTTCA AATCTTAAGT TTTAAGTTTT CTTTAAGGAA AGTATATTAT TCTGAAGGAC 1260

TCTAAAATTT CGCAGCCATT TATTAGTAAT TGCTACAGAA TTCCTAGTCA TTACTAGAAA 1320

TGGACTAGTT TCTTTGAATA ATAGAACTGC ATAATTCTCC TATTCTAGAA GGGGAGGACC 1380

AGTATTTCTT TTATGATAGG ACTAGATTGT GGTATAATAG AGAGAATAAG TTTTTTTAGT 1440

AAGACAAAGG AGAAAATAGA TGATTTATGC AGGAATTCTT GCCGGTGGAA CTGGCACACG 1500

CATGGGGATC AGTAACTTGC CAAAACAATT TTTAGAGCTA GGTGATCGAC CTATTTTGAT 1560

TCATACAATT GAAAAATTTG TCTTGGAGCC AAGTATTGAA AAAATTGTAG TTGGTGTTCA 1620

TGGAGACTGG GTTTCTCATG CAGAAGATCT TGTAGATAAA TATCTTCCTC TTTATAAGGA 1680

ACGTATCATC ATTACAAAGG GTGGTGCTGA CCGCAATACA AGTATTAAGA ACATCATTGA 1740

AGCCATTGAT GCTTATCGTC CGCTTACTCC AGAGGATATC GTTGTTACCC ACGATTCTGT 1800

TCGTCCATTT ATTACACTTC GCATGATTCA GGACAATATC CAACTTGCCC AAAATCATGA 1860

CGCAGTGGAC ACAGTGGTAG AAGCGGTTGA TACTATCGTT GAAAGTACCA ATGGTCAATT 1920

TATTACAGAT ATTCCAAATC GTGCTCACCT TTATCAAGGA CAAACACCTC AAACATTCCG 1980

TTGCAAGGAC TTCATGGACC TTTATGGATC TCTTTCTGAT GAAGAGAAGG AAATCTTGAC 2040

AGATGCATGT AAAATCTTTG TGATCAAAGG AAAAGATGTG GCTTTGGCCA AAGGTGAATA 2100

CTCAAATCTG AAGATTACAA CCGTAACAGA TTTGAAGATT GCAAAAAGTA TGATTGAGAA 2160

AGACTAGTAA AATGATTAAT CAAATTTATC AACTAACTAA GCCTAAGTTT ATCAATGTCA 2220

AATATCAGGA AGAGGCTATT GACCAAGAGA ATCATATCCT TATCCGTCCC AACTACATGG 2280

CTGTCTGTCA TGCGGATCAG CGTTACTATC AGGGAAAACG TGATCCCAAG ATTTTGAATA 2340

AAAAGCTTCC AATGGCAATG ATTCACGAGT CATGTGGAAC CGTCATTTCT GACCCGACCG 2400

GAACCTACGA GGTTGGTCAA AAAGTTGTCA TGATTCCCAA TCAGTCTCCT ATGCAGAGTG 2460

ATGAAGAATT CTATGAAAAC TACATGACAG GGACCCATTT CTTGTCTAGT GGATTTGATG 2520

GCTTTATGAG AGAGTTTGTT TCTCTCCCTA AAGATCGTGT GGTGGCTTAT GATGCTATTG 2580

AAGATACGGT TGCAGCCATT ACAGAGTTTG TCAGTGTGGG CATGCACGCT ATGAATCGTC 2640

TATTGACTCT TGCTCATAGC AAGCGGGAGC GGATCGCCGT TATTGGAGAT GGAAGTTTAG 2700

CTTTTGTGGT TGCCAATATT ATCAACTATA CTTTGCCAGA AGCAGAGATT GTGGTTATTG 2760

GTCGTCATTG GGAAAAGTTG GAACTCTTCT CATTTGCCAA AGAATGCTAT ATTACGGATA 2820

ATATTCCTGA AGATTTGGCC TTTGACCATG CTTTTGAATG TTGTGGTGGT GATGGTACTG 2880

GACCAGCTAT TAATGACTTG ATTCGCTACA TTCGTCCTCA GGGAACGATT CTCATGATGG 2940

GAGTTAGCGA ATATAAAGTC AATCTCAATA CTCGCGATGC CTTAGAAAAG GGCTTGATTT 3000

TGGTTGGGTC ATCTCGTTCT GGTCGCATTG ATTTTGAAAA TGCTATCCAA ATGATGGAAG 3060

TCAAGAAATT TGCCAATCGT CTTAAAAATA TCCTTTATCT AGAAGAACCT GTAAGAGAAA 3120

TTAAAGATAT TCATCGTGTC TTTGCAACCG ATTTAAACAC AGCCTTTAAA ACAGTGTTTA 3180

AGTGGGAAGT ATAAGTACTG GAGGTTAATT GTGGAGAAAA TCATTAAAGA AAAAATTTCT 3240

TCCTTACTTA GTCAAGAAGA GGAAGTCCTC AGTGTTGAAC AACTGGGTGG AATGACCAAT 3300

CAAAACTATT TGGCCAAAAC AACAAATAAG CAATACATTG TTAAATTCTT TGGTAAAGGG 3360

ACAGAAAAGC TTATCAATCG ACAAGATGAA AAGTACAATC TTGAACTACT AAAGGATTTA 3420

GGCTTAGATG TAAAAAATTA TCTTTTTGAT ATTGAAGCTG GTATCAAAGT AAATGAGTAT 3480

ATCGAATCTG CGATTACGCT TGATTCAACG TCAATCAAGA CCAAGTTCGA CAAAATTACT 3540

CCAATATTAC AAACTATTCA TACGTCTGCT AAGGAATTAA GAGGAGAATT TGCTCCTTTT 3600

GAAGAAATCA AAAAATACGA ATCCTTGATT GAAGAACAAA TTCCTTATGC CAACTATGAA 3660

TCTGTTAGAA ATGCAGTCTT CTCCTTAGAG AAAAGACTGG CTGACTTAGG TGTTGACAGA 3720

AAATCTTGTC ATATCGATTT GGTGCCTGAA AACTTTATCG AATCACCTCA AGGACGACTT 3780

TATTTGATTG ACTGGGAATA TTCATCAATG AATGATCCAA TGTGGGATTT GGCTGCCCTC 3840

TTTTTAGAGT CTGAATTCAC TTCCCAAGAG GAAGAAACTT TCTTATCTCA CTATGAGAGT 3900

GACCAAACAC CGGTTTCTCA TGAAAAGATT GCTATTTATA AAATTTTACA AGATACTATT 3960

TGGAGTCTAT GGACTGTCTA TAAGGAAGAG CAAGGTGAAG ATTTTGGTGA CTATGGTGTG 4020

AATCGTTACC AAAGAGCTAT TAAAGGTTTG GCTTCTTATG GAGGTTCAGA TGAAAAGTAA 4080

AAACGGAGTT CCTTTTGGCC TTCTCTCAGG TATTTTCTGG GGCTTGGGTC TAACGGTTAG 4140

TGCTTATATC TTTTCGATTT TTACAGATTT GTCACCCTTT GTGGTGGCTG CAACTCATGA 4200

TTTTTTGAGC ATCTTTATCT TACTAGCTTT TCTCTTGGTA AAAGAAGGGA AAGTTCGCCT 4260

CTCAATTTTC TTAAATATTC GCAATGTCAG TGTTATCATC GGAGCCTTGC TAGCAGGCCC 4320

TATCGGTATG CAGGCCAATC TTTATGCAGT TAAGTATATC GGAAGTTCTT TAGCTTCATC 4380

TGTATCGGCT ATTTACCCTG CGATTTCAGT TCTATTGGCT TTCTTCTTTT TGAAGCACAA 4440

GATTTCGAAA AATACTGTAT TTGGGATTGT CTTGATTATT GGAGGGATTA TTGCTCAGAC 4500

CTATAAGGTT GAACAGGTTA ATTCTTTCTA CATTGGGATT CTTTGTGCTT TGGTTTGTGC 4560

TATTGCATGG GGAAGTGAGA GTGTTCTTAG CTCTTTTGCC ATGGAAAGTG AATTGAGTGA 4620

AATCGAAGCC CTCTTAATCC GTCAAGTAAC TTCGTTCTTG TCCTATCTTG TGATTGTGCT 4680

CTTCTCTCAT CAGTCATTTA CTGCAGTAGC CAATGGACAA TTGCTAGGTC TCATGATTGT 4740

TTTTGCAGCC TTTGATATGA TTTCCTACTT GGCTTATTAT ATCGCTATCA ATCGCTTGCA 4800

ACCAGCCAAG GCTACAGGCT TGAACGTGAG CTATGTAGTA TGGACGGTCT TGTTTGCAGT 4860

TGTTTTCTTG GGTGCACCGC TAGATATGCT GACCATTATG ACGTCACTTG TCGTCATTGC 4920

TGGAGTTTAT ATTATTATTA AAGAATAAAG GAGATTCGTG TGAAAGCCAT TATCTTAGCA 4980

GCGGGATTGG GAACTCGCTT GCGTCCTATG ACTGAAAATA CCCCTAAAGC CTTGGTTCAG 5040

GTTAATCAAA AACCTTTGAT TGAGTACCAA ATTGAGTTTC TCAAAGAAAA AGGAATCAAT 5100

GACATCATCA TCATTGTTGG TTATCTTAAA GAACAATTCG ATTACTTGAA AGAGAAATAC 5160

GGTGTTCGTC TCGTTTTCAA TGATAAATAC GCTGACTACA ATAACTTTTA CTCTCTCTAT 5220

CTTGTAAAAG AAGAATTGGC CAACAGCTAT GTTATTGATG CTGACAATTA TCTCTTTAAA 5280

AATATGTTCC GCAATGATTT GACACGTTCG ACTTATTTTA GTGTTTATCG TGAAGATTGT 5340

ACCAACGAAT GGTTCTTGGT TTATGGAGAT GACTACAAGG TTCAAGACAT TATTGTTGAT 5400

AGCAAGGCAG GTCGCATCCT TAGTGGTGTA TCCTTCTGGG ATGCTCCAAC TGCAGAAAAG 5460

ATTGTCAGCT TTATCGACAA GGCTTATGTA AGTGGTGAAT TTGTTGATCT CTATTGGGAC 5520

AATATGGTTA AGGATAATAT CAAAGAGCTA GATGTCTATG TTGAAGAATT AGAAGGCAAT 5580

AGCATTTATG AGATCGATAG TGTCCAAGAC TATCGTAAAT TAGAAGAAAT TCTTAAAAAC 5640

GAAAATTAAA GATTCCAACA TCTGACAAAA TAGTCGGATG TTTTTTGATT TTTTACGAAC 5700

TTTTACGAAT AGATAGATGA GTAGAAAAAG AAATGGAGTT ATTTATGAAA ATCACAAACT 5760

ATGAAATCTA TAAGTTAAAA AAATCAGGTT TGACCAATCA ACAGATTTTG AAAGTGCTAG 5820

AATACGGTGA AAATGTTGAT CAGGAGCTTT TGTTGGGTGA TATTGCAGAT ATCTCAGGTT 5880

GCCGTAATCC AGCCGTTTTT ATGGAACGTT ATTTTCAGAT AGACGATGCG CATTTGTCGA 5940

AAGAGTTTCA AAAATTTCCA TCTTTCTCTA TTTTAGATGA CTGTTATCCT TGGGATTTGA 6000

GTGAAATATA TGATGCGCCT GTACTTTTAT TTTACAAGGG AAATCTTGAC CTCCTGAAAT 6060

TCCCGAAGGT AGCGGTCGTG GGCAGTCGTG CTTGTAGCAA ACAGGGAGCT AAGTCAGTTG 6120

AAAAAGTCAT TCAAGGCTTG GAAAATGAAC TGGTTATTGT CAGTGGTCTG GCCAAGGGCA 6180

TTGACACAGC AGCTCATATG GCAGCTCTTC AGAATGGCGG AAAAACCATT GCAGTGATTG 6240

GAACAGGACT GGATGTGTTT TATCCTAAAG CCAATAAACG CTTGCAAGAC TACATCGGCA 6300

ATGACCATCT GGTTCTAAGT GAATATGGAC CTGGTGAACA ACCTCTGAAA TTTCATTTTC 6360

CTGCCCGTAA TCGCATCATT GCTGGACTTT GTCGTGGTGT GATTGTAGCA GAGGCTAAGA 6420

TGCGTTCAGG TAGTCTCATT ACGTGTGAGC GAGCAATGGA AGAAGGACGC GATGTCTTTG 6480

CTATTCCTGG TAGCATTTTA GATGGACTAT CAGACGGTTG CCATCATTTG ATTCAAGAAG 6540

GAGCAAAATT GGTCACCAGT GGGCAAGATG TTCTTGCGGA ATTTGAATTT TAAAAATGAC 6600

CTAAGCTAGA ATTCTAAGAA AAAATCAATT TTAAGAGAAA ATGAACCCAA CATTTCCATA 6660

ATAAAACGCA TATTAGCAAG TTTTTAACAC TTGATAATAT GCGTTTTTTC TAAGTGGATT 6720

AGTAGAGTAG AGGATTTTTC TCATATAATA CTCTTCGAAA ATCTCTTCAA ACTACGTCAG 6780

CTTCCATCTG CAACCTCAAA ACAGTATTTT GAGCgaCTtC GTCAGTCTTA TCTACAACCT 6840

CAAAGCAGTG CTTTGAGCAA CCTGTGGCTA GCTTCCTAGT TTGCGCTTTG ATTTTCATTG 6900

AGTATAAGGG AAAGTATAGT GAATTGAAAT AAGATGTGAA CAACTCTATC AGGAAAGTCA 6960

AATTAATTTA TAGAAATATT TTAGCAGCCA AGGTGTACTG TTATAGATTC AATTACACTA 7020

TAATTTAGTG TAATTGAGAA AGGAGAAATG ATTGTGATTG ATGTTGGCTA GGTTATGTTC 7080

AATGATTCCT ACCGTCTCAA ATCTTGTCAG TAAGGAAAAA TAAATTCTTC AAAAGTAGAG 7140

ATTACAAGGC TTGTTTAAGA AAGAATTCAA AGACCTTGAC AAATAAAAAT AAAATGGTTA 7200

TTATAAAAAA TGGTCTGAAA TAGATGATGA TACTTTTCGA AAATCTCTTC AAATACGTCA 7260

GCTCAGCTTT GCCTTGCTGT GTTTTGAGCA AGCTACGGTT AGCTTCCGAG TTTGATTTTC 7320

ATTTACTAGA AATGAAACTG ATGAGAGATA TCAGTAGACA TTTGAGTCAG GATATTATGG 7380

AAAATGATAA AAAGAGCTCG TGAGATTGGC ATATCAGACT ACTAAAGTAT TGAGTTTGTT 7440

AGGATTTTAG CGACTAGTTA GCTGGGAAAG GAAGATATTT GTGACAAATA ATAAACTGTA 7500

TTCGTTGATA GAATTTAGAA ATAAAATATA TGAAGAATTA GAACTTTCCA GAAGTGATTT 7560

AGCGATTTTA CTATGTGCCA TGCTTATCGC CTCTATCGGA TTAAATATGG ATTCGACTCC 7620

CGTGATTATT GGAGCCATGT TAATCTCTCC TTTGATGACA CCTATTCTGG GAGTGGGGCT 7680

CTCTCTAGCT ATATTTGATT TTAAATTGTT AAGAAAATCT TTTAAAATAT TAGCTATTCA 7740

AATTCTTGCC AGTCTAATAG CTTCAACACT TTATTTTTAT CTTTCTCCCA TTTCGTATGC 7800

TAGTTCGGAG ATTGTTGCTA GAACCTCTCC GACTATTTGG GATGTTCTCA TTGCTTTTGT 7860

AGGAGGGATA GCAGGTATCA TTGGTGCTAG GAAAAAAGAG AC 7902

(2) INFORMATION FOR SEQ ID NO: 113:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18627 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

GAAGTTGAAA TGGCCAGCTG ATGAGCAATA TCGGTCATAG AAATCTTCTC AATCAACTTT 60

TGCGCAATTT TTTGGTTGAT AATACGAGGA ATTTGGTGAT TTTTCTTGAC GATAGAAGTT 120

TCAGCGACCA TCATTTTTGA ACAGTGATAG CACTTGAAAC GACGCTTTCT AAGTAGAATT 180

CTAGTAGGCA TACCAGTTGT CTCAAGGTAA GGAATCTTAG ACGGTTTTTG AAAGTCATAT 240

TTCTTCAATT GGTTTCCGCA CTCAGGGCAA GATGGGGCGT CGTAGTCCAG TTTGGCGATG 300

ATTTCCTTGT GTGTATCTTT ATTGATGATG TCTAAAATCT GGATATTAGG GTCTTTAATG 360

TCTAGTAATT TTGTGATAAA ATGTAATTGT TCCATATGAA TCTTTCTAAT GAGTTGTTTG 420

GTCGCTTTTC ATTATAGGTC ATATGGGACT TTTTTTCTAC AATAAAATAG GCTCCATAAT 480

ATCTATAAGG GATTTACCCA CTACAAATAT TATAGAGCCA AAAATCCTTT GTTTACTAAA 540

CAAGGGATTT TTCTTTTGTC TCTGCTCCTT TTTTGATATA ATAGTTCTAT GTTAAAATCA 600

GAAAAACAAT CACGTTATCA AATGTTAAAT GAAGAATTGT CCTTCCTATT GGAAGGCGAA 660

ACCAATGTTT TGGCTAATCT TTCCAACGCC AGTGCTCTCA TAAAATCACG TTTTCCTAAT 720

ACCGTATTTG CAGGCTTTTA TTTGTTCGAT GGAAAGGAAT TGGTTTTAGG CCCCTTCCAA 780

GGAGGTGTTT CCTGCATCCG TATTGCACTA GGCAAGGGTG TTTGTGGTGA GGCAGCTCAC 840

TTTCAGGAAA CTGTTATTGT TGGAGATGTG ACGACCTATC TCAACTATAT TTCTTGTGAT 900

AGTCTAGCTA AAAGTGAAAT TGTGGTGCCG ATGATGAAGA ATGGTCAGTT ACTTGGAGTT 960

CTGGATCTGG ATTCTTCAGA GATTGAGGAT TACGATGCTA TGGATCGAGA TTATTTGGAA 1020

CAATTTGTCG CTATTTTGCT TGAAAAGACA GCATGGGACT TTACGATGTT TGAGGAAAAA 1080

TCTTAATGTA TCAAGCACTT TATCGAAAAT ATAGAAGTCA AAACTTCTCC CAGTTAGTTG 1140

GTCAAGAAGT TGTGGCTAAG ACTCTTAAAC AAGCGGTGGA GCAAGAGAAA ATAAGTCACG 1200

820

CTTATCTTTT TTCTGGTCCT CGTGGAACGG GAAAAACCAG TGTTGCTAAA ATCTTTGCCA 1260

AGGCTATGAA CTGTCCCAAT CAAGTGGGTG GCGAACCTTG CAATAACTGC TATATTTGTC 1320

AAGCAGTGAC GGACGGTAGT TTAGAAGATG TCATTGAAAT GGATGCAGCT TCTAATAATG 1380

GGGTAGATGA AATTCGCGAA ATTCGTGATA AATCTACCTA TGCGCCTAGC CTTGCTCGTT 1440

ATAAGGTTTA TATCATAGAT GAGGTTCACA TGCTGTCTAC AGGGGCTTTT AATGCCCTCC 1500

TAAAGACGCT GGAAGAACCA ACACAGAATG TAGTCTTTAT TTTGGCCACT ACTGAATTGC 1560

ACAAGATTCC TGCTACTATT CTATCCCGTG TGCAACGTTT TGAGTTTAAA TCAATTAAGA 1620

CACAGGATAT TAAGGAACAT ATTCACTATA TCTTAGAAAA AGAAAATATC AGTTCTGAAC 1680

CAGAGGCTGT GGAAATCATT GCCAGACGGG CGGAAGGTGG AATGCGGGAC GCCTTGTCTA 1740

TTTTGGATCA AGCCCTGAGT TTGACACAGG GAAATGAGCT GACGACTGCT ATCTCTGAAG 1800

AAATTACTGG CACCATTAGC CTATCAGCCT TGGATGATTA TGTGGCGGCC TTGTCTCAAC 1860

AGGATGTTCC CAAAGCTTTG TCTTGCTTGA ATCTTCTTTT TGACAATGGT AAGAGCATGA 1920

CTCGTTTTGT GACCGATCTT TTGCACTATT TAAGAGACTT GTTAATTGTT CAAACAGGGG 1980

GAGCAAATAC TCATCATAGT TCAGTCTTTG TAGAAAATTT GGCACTTCCT CAAAAAAATC 2040

TGTTTGAAAT GATTCGCTTA GCAACAGTGA GTTTAGCAGA TATTAAGTCT AGTTTGCAAC 2100

CCAAGATTTA TGCTGAAATG ATGACCGTCC GTTTGGCGGA AATCAAGTCC GAACCAGCTC 2160

TATCAGGAGC GGTTGAAAAT GAAATTGCTA CGCTGAGACA GGAAGTTGCC CGTCTCAAAC 2220

AAGAGCTTTC TAATGTAGGT GCGGTTCCTA AACAAGTTGC ACCAGCTCCT AGTCGACCAG 2280

CTACGGGCAA AACAGTCTAT CGTGTCGATC GCAATAAAGT GCAATCTATC TTACAAGAGG 2340

CCGTCGAAAA TCCTGATTTA GCACGTCAAA ATTTAATTCG TTTGCAGAAT GCCTGGGGAG 2400

AGGTAATTGA AAGTCTAGGT GGGCCGGACA AGGCTCTGCT AGTTGGTTCT CAACCGGTTG 2460

CTGCCAATGA ACACCATGCT ATTCTTGCTT TTGAGTCTAA CTTCAATGCT GGTCAAACTA 2520

TGAAACGAGA CAATCTCAAT ACCATGTTTG GTAATATCCT CAGTCAGGCG GCAGGTTTTT 2580

CACCTGAGAT TTTAGCTATT TCCATGGAGG AATGGAAAGA AGTTCGCGCA GCCTTTTCAG 2640

CCAAAGCCAA ATCTTCTCAA ACTGAAAAAG AAGTAGAAGA AAGCCTGATT CCAGAAGGAT 2700

TTGAATTTTT GGCTGATAAA GTGAAGGTAG AGGAAGACTA AAGAAAGATT TCATGATACA 2760

ATAAGTTTAT GAATAAACAA CAATTTATTA TTATGGCGCT GTTTACAGCT GCTGAGACCT 2820

ATTTTTTCAA TGAAGCCTGG ATGACTGGCC GCTATATTAT GGCAGCCTTT TGGGCAATTT 2880

TACTCTTTAG AAATTTCCGA GTCAGTTATG TGATGGGCAA AATCGTTGAT GTCATCGATC 2940

AGCATTTTAA TAGGAAAGAC TAGCCCTCAG CTTCCAGACA AAATCAAAGC CTTTTAGGCT 3000

TTTTTTTGTT ATACTAGAAA AGTATATTTA TAGAATTTTT GCTCTATTTC TGGGGAAATC 3060

AGACGTTTTT CTAGTAAGTA CTGTAAAAGT TTTGAAAAAG AAAGGAACTA TCATGTCAGT 3120

ATTAGAGATC AAAGATCTTC ACGTTGAGAT TGAAGGAAAA GAAATTTTAA AAGGGGTTAA 3180

CCTGACCCTG AAAACAGGAG AAATTGCCGC TATCATGGGA CCAAATGGTA CAGGTAAATC 3240

GACTCTTTCT GCCGCTATCA TGGGAAATCC AAACTATGAA GTAACTAAAG GTGAAGTTTT 3300

GTTTGATGGC GTAAACATCC TTGAGTTGGA AGTGGATGAG CGTGCGCGTA TGGGACTTTT 3360

CCTTGCTATG CAATACCCAT CAGAAATCCC TGGAATTACC AATGCTGAGT TTCTTCGTGC 3420

CGCTATGAAT GCGGGTAAAG AAGATGATGA GAAGATTTCA GTTCGTGAGT TTATTACTAA 3480

GCTAGATGAA AAAATGGAAT TGCTCAACAT GAAAGAAGAA ATGGCAGAGC GTTACCTCAA 3540

CGAAGGCTTC TCTGGTGGTG AGAAAAAACG CAATGAAATT CTTCAACTTT TGATGTTGGA 3600

GCCAACATTT GCTCTTTTGG ACGAGATTGA CTCAGGTCTT GATATTGACG CTCTTAAAGT 3660

TGTGTCTAAA GGTGTCAATG CCATGCGTGG TGAAGGTTTT GGTGCTATGA TCATCACTCA 3720

CTACCAACGT CTTTTGAACT ATATCACACC TGATGTGGTA CACGTGATGA TGGAAGGTCG 3780

TGTTGTCCTT TCTGGTGGTC CAGAATTGGC TGCGCGTTTG GAACGTGAAG GATACGCAAA 3840

ATTAGCTGAA GAACTTGGCT ACGACTACAA GGAAGAATTG TAATTCCCTC GTATCTTTTA 3900

GGAGAAGTAA ATGACTAGAG AAAATATTAA ACTTTTTTCA GAAATGCACG CTGAACCAAG 3960

CTGGTTGGCT GATCTCCGTC AAAAAGCTTT TGACAAGATT GAGACTTTGG AATTACCAGT 4020

TATTGAGTGT GTCAAATTCC ACCGTTGGAA TCTGGGTGAT GGAACGATTA CAGAAAATGA 4080

GCCATCAGCA AATGTTCCAG ATTTCACAGC TTTAGATCAT CACTTGAAGT TGGTGCAAGT 4140

AGGAACTCAA ACTGTTTTCG AACAAACTCC AGTTGAGTTA GCTGAACAGG GTGTTGTCTT 4200

CACAGACTTT CACTCAGCTT TAGAAGAAAT TCCAGAGCTG ATCGAAGAAT TCTTCATGTC 4260

ATCTGTTAAG TATGATGATG ACAAGTTGGC GGCTTACCAC ACAGCTTACT TTAACAGTGG 4320

TGCTGTACTC TATATTCCAG ATAACGTAGA AATCACAGAG CCAATTGAAG GAATTTTCTA 4380

CCAAGATAGC GATAGCAATG TGCCGTTTAA CAAGCATATT ATGATTATCG TTGGTAAAAA 4440

TTCTAAGATT AGTTATCTGG AGCGTTTAGA GTCACGCGGT GAAGGAAGTG ACAAAGCAAC 4500

TGCCAATATC ACAGTGGAAG TGATTGCACG TTCTGGTGCG CAAGTCAAGT TTGCTGCTAT 4560

CGACCGTCTA GGTGAAAACG TCACTGCCTA CATTAGCCGT CGTGGTAAAT TAGGCAACGA 4620

TGCAAGTATT GACTGGGCTA TCGGTGTCAT GAACGAAGGA AATGTCGTTG CTGATTTTGA 4680

TAGTGACTTG ATTGGTAATG GTAGCCATGC TGACCTCAAG GTTGTAGCTC TTTCAAGTGG 4740

TCGTCAGGTA CAAGGGATTG ATACTCGTGT AACTAACTAT GGCTGCAACT CAATCGGAAA 4800

CATTCTACAA CATGGGGTTA TCCTTGAAAA AGCAACTTTG ACTTTCAATG GTATCGGCCA 4860

CATCATCAAG GGTGCTAAGG GAGCAGATGC GCAACAAGAG AGCCGTGTTC TCATGCTTTC 4920

AGACCAAGCG CGTTCAGATG CTAACCCAAT TCTTTTGATT GATGAAAATG ACGTAACTGC 4980

AGGCCATGCA GCCTCTATTG GTCAGGTAGA TCCAGAAGAT ATGTACTACC TCATGAGTCG 5040

TGGCTTGGAT AAGGCAACTG CAGAGCGTTT GGTTGTTCGT GGTTTCCTTG GATCTGTTAT 5100

CGTGGAGATT CCAGTCAAGG AAGTTCGTGA TGAAATGATT GCAACTATCG AAGAGAAATT 5160

GTCAAAACGC TAAGGGGCAG CCTATGTTAG ATGTAGAAGC GATTCGCAAG GATTTTCCAA 5220

TTTTAGATCA GATTGTCAAT GATGAACCTC TGGTCTATCT GGACAATGCT GCGACGACAC 5280

AAAAACCACT AGTAGTTCTG AAAGCTATTA ACAGCTACTA TGAGCAGGAC AATGCCAATG 5340

TTCACCGTGG TGTCCATACC TTAGCGGAAC GAGCGACAGC TTCTTATGAA GCTGCTCGTG 5400

AAACCATTCG TAAGTTTATT AATGCAGGCT CTACAAAGGA AGTTCTCTTT ACCAGAGGAA 5460

CGACAACCAG CCTTAACTGG GTGGCACGCT TTGCTGAGGA AATTCTCACT GAGGGAGACC 5520

AGGTCTTGAT TTCAGTAATG GAACACCATT CTAATATCAT TCCATGGCAG GAAGCTTGTC 5580

GAAAGACTGG AGCAGAGCTT GTCTATGTCT ATCTTAAAGA CGGTGCCTTG GATATGGAGG 5640

ATTTGCGAGC TAAATTGACT GATAAGGTTA AATTTGTTTC CCTAGCTCAT GCCTCCAATG 5700

TTCTTGGTGT GGTCAATCCG ATCAAGGAAA TCACTCAATT AGCCCACCAA GTTGGGGCAA 5760

TTATGGTAGT GGATGGTGCT CAATCTACAC CTCATATGAA GATTGATGTC CAGGACTTGG 5820

ATCTGGACTT TTTCGCCTTT TCGGGTCACA AGATGGCTGG TCCGACTGGT ATCGGTGTCC 5880

TTTACGGCAA AGAAAAGTAT CTTGAGCAAA TGTCTCCAGT AGAATTTGGC GGCGAGATGA 5940

TTGATTTTGT CTACGAGCAA TTTGCTAGTT GGAAGGAATT GCCTTGGAAA TTTGAGGCTG 6000

GAACGCCAAA TATGGCAGGA GCTATTGGAC TTGCGACTGC AGTTGATTAT CTGGAAAAGA 6060

TTGGTATGGA TGCCGTTGAA GCTCATGAAC AGGAATTGAT TGCGTACGTC TATCCAAAAC 6120

TGCAGGCAAT TGAGGGATTG ACCATTTACG GTTCTCAGGA TTTGGCTCAA CGTTCGGGTG 6180

TTATTGCCTT TAACCTAGGT GATCTCCATC CTCACGATCT TGCGACGGCT CTGGATTATG 6240

AAGGAGTGGC TGTTCGTGCT GGTCACCATT GTGCGCAACC CTTGCTTCAG TATTTGGAAG 6300

TCCCAGCAAC AGCTCGTGCA AGTTTTTATA TCTACAATAC CAAGGCAGAT TGCGACAAAC 6360

TAGTCGATGC CCTACAAAAG ACAAAGGAGT TTTTCAATGG CACTTTCTAA ACTAGATAGC 6420

CTTTATATGG CAGTGGTAGC AGACCATTCG AAAAATCCAC ATCACCAAGG GAAGTTAGAA 6480

GATGCTGAGC AAATCAGTCT CAACAATCCG ACTTGTGGGG ATGTCATCAA CCTCTCTGTC 6540

AAGTTTGATG CAGAGGACCG TTTGGAAGAT ATTGCTTTTC TAAATTCAGG ATGCACGATT 6600

TCAACTGCTT CTGCTAGTAT GATGACAGAT GCCGTTTTAG GAAAAACCAA ACAAGAAATT 6660

TTAGAACTGG CGACTATTTT TTCTGAAATG GTTCAAGGGC AAAAAGATGA GCGTCAAGAC 6720

CAACTTGGAG ACGCGGCATT CTTGTCAGGT GTTGCCAAAT TCCCTCAAAG AATCAAGTGT 6780

GCAACCCTAG CTTGGAATGC CCTTAAGAAA ACAATTGAAA ATCAAGAAAA ACAGTAAGAC 6840

AAGTTTCTTT TGTCTTATGA ATTATTAGAA ATGAAGAAAG AAAGGATACT ATGGCTGAAG 6900

AAAGAGTAGA ACCAAAACCA ATTGACCTTG GTGAATATAA ATTTGGTTTC CATGACGATG 6960

TAGAGCCTGT CTTATCGACA GGAAAAGGAC TCAACGAAGG TGTTATTCGT GAATTATCTG 7020

CTGCTAAGGG TGAGCCTGAG TGGATGTTGG AGTTCCGTTT GAAGTCTTAT GAAACCTTCA 7080

AAAAAATGCC CATGCAAACT TGGGGAGCAG ACTTGTCAGA GATTGACTTT GATGACTTAA 7140

TCTACTACCA AAAACCATCT GACAAACCAG CCCGTTCTTG GGATGATGTA CCTGAAAAGA 7200

TTAAAGAAAC CTTTGAACGT ATCGGGATTC CAGAAGCTGA ACGTGCTTAT TTAGCAGGGG 7260

CTTCTGCCCA GTACGAGTCA GAAGTGGTTT ACCACAACAT GAAGGAAGAG TTCCAAAAAT 7320

TAGGTATTAT CTTTACAGAT ACAGATTCCG CACTCAAGGA ATACCCAGAC TTATTTAAAC 7380

AATACTTTGC GAAGTTGGTA CCGCCGACAG ATAACAAGTT GGCAGCCCTC AACTCAGCAG 7440

TATGGTCGGG TGGAACTTTT ATCTACGTGC CAAAAGGTGT CAAGGTAGAT ATTCCACTTC 7500

AAACTTATTT CCGTATCAAT AACGAAAATA TAGGTCAGTT CGAACGTACC TTGATTATCG 7560

TTGATGAGGG AGCAAGCGTC TACTACGTAG AAGGATGTAC AGCACCAACA TATTCAAGCA 7620

ATAGCTTACA CGCTGCCATT GTAGAAATTT TTGCTTTGGA CGGAGCTTAT ATGCGTTATA 7680

CAACTATCCA AAACTGGTCT GATAACGTCT ATAACTTGGT AACAAAGCGT GCTAAGGCTC 7740

AAAAGGATGC CACTGTTGAG TGGATTGATG GAAACTTGGG TGCCAAAACG ACTATGAAAT 7800

ATCCATCTGT TTACCTTGAT GGAGAAGGAG CGCGTGGTAC CATGCTCTCT ATCGCCTTTG 7860

CTAATGCAGG GCAACACCAA GACACGGGTG CTAAGATGAT TCACAATGCT CCACATACCA 7920

GCTCGTCTAT TGTGTCTAAA TCCATCGCTA AAGGTGGAGG AAAGGTTGAC TACCGTGGAC 7980

AAGTCACCTT TAACAAGAAC TCTAAGAAAT CTGTTTCCCA CATTGAATGT GATACCATTA 8040

TCATGGATGA CTTGTCAGCA TCAGATACTA TTCCATTTAA TGAAATTCAC AACTCGCAAG 8100

TGGCTTTGGA ACACGAAGCC AAAGTATCTA AGATTTCAGA AGAGCAATTG TATTATCTCA 8160

TGAGCCGTGG ATTGTCAGAA TCTGAGGCAA CTGAAATGAT TGTCATGGGA TTTGTAGAAC 8220

CCTTTACAAA AGAACTTCCA ATGGAATACG CAGTTGAGCT GAACCGCTTG ATTAGCTATG 8280

AAATGGAGGG ATCAGTTGGA TAAAATTTGA TTTTATACTC TTCGAAAATC TCTTCAAACC 8340

ACGTCAGCAT CGCCTTACCG TATGTATGGT TwCTGAtTCG TCAGTTTCAT CTACAACCTC 8400

AAAACAGTGT TTTGAGCAAC tGCGGCTAGC TTCCTAGTTT GTTCTTTGAT TTTGAGTATT 8460

AGATTTACTC AAAATCAAGG ATTTTGAAGA TGAACTTGTA TCAAAAAATC GCGGTTTAAA 8520

ATCGCGATTT TTTATAATTT CTCGTTAACA AAGCGGACAA ACTGATTCCA CCAAACTTTT 8580

AAGAAGAAGG CTTTTTCAAT TTTCTTGTCT GCTACCATTT CGAAACTAGG GCGCTCTGTG 8640

GTGATGTAAC CTTGACCAAT CAAGTCCTTG TCTTCATAAG TCAAATGGCC AACCACTGTT 8700

CCAGCTTCAA GTGGTGCTGG GATTGCTTTG GAATCAGGTG TGAATTGAAC AGATTGGGAA 8760

GATTGATTCC CAACACGTTC GATTAGATAG ATATCCTCTG GAGCCACTGC AGTTACTGTA 8820

TCTTCTTTTC CATCTTGTAC AGGGGCTTTG CTATCTTGAT AGGCATCGCC TTGTTGAACG 8880

ATTTTGCGAA GTGTAAATGT AGAAGAAATA TAATCCATTA GGGAAGATGT AGCTGTAAAT 8940

CGAGCGTAAG GATTATTGTC TTGATGATCT GCATTTAAAA CAACTGTGAT GACTCTCATG 9000

CCTTTTTCGA CAGTAGTACC AACAAAAGAC TCTCCAGCCT TATCTGTTGT TCCTGTTTTT 9060

AGCCCATCAA AACCACCACG GTAAGCAGGC ATACCTTCTA ACATGTAGTT GGTTGAAGTG 9120

ATTGTCATCC CAGCAAAAGT AGAAGAAGGT TTTTTGGTGA TTTCTAAGAC TTGTGGGTAT 9180

TTTTTGATGA GGTTGCGAGC AACGATAGCG ACATCATAAG CACTAAGCTT ATTTTCCTCA 9240

TCTTTTTTAG AACCTGGGTA AATGTTATCC CCTAGAGTTT CATTGTTAAG ACCTGTCGTA 9300

TTGACAACAG TGGCATCCTG AATTCCCCAT TCCAAGAGTT TTGCCCGCAT CATATCGACG 9360

AAATCTTTTT CTGAGCCAGC AATTTTCTCA GCTAGGGCAA TAGCGGCGCT GTTGGCACTA 9420

GATACCAGAG TTGCTTCAAG CAACTCTTCG ACAGTATAAT TACGGGCCTC CATAGGAATA 9480

TTACTGGCTT CAGAATTTGT CGTCAATTGA TAAGGATAAT CAGAAATATC TACAGGAGTG 9540

GAGAGGGTAA TACTTCCGTT TTCCAAAGCT TCATAGACCA GATAAACAGT AATCAATTTT 9600

GTTATGGAAG CAATTTCGAC AGGTTGCGTT GCATCCTTCT CATAGAGAAT TTTACCAGTA 9660

TTTGCCTCAA CAGCAATCGC ATGTTTAGCG GCAATGGTAA AATCTTGAGC AACAGCAGTA 9720

GAAGCACCCC CTAAAAGAGA GACAGTTAAC AAAGTTAAAA ATATTTTTTT CATAGTAGTC 9780

TTATTCTATC ATAAAGAAAA AAAATATTCT TGCTTTAATA ATTCATCTGT TAAGCTTTTT 9840

GAAAATATGG TAAAATAAAG TAAGGGAGGT AACTCATGTT TCGTAGAAAT AAATTATTTT 9900

TTTGGACCAC AGAAATTTTA CTCTTAACCA TCATCTTTTA CCTATGGAGA CAGATGGGGT 9960

CTTTGATTAA CCCTTTTGTT AGCGTGCTTA ATACAATTAT GATTCCATTT TTATTAGGGG 10020

GCTTTTTTTA TTATTTGACA AACCCTATTG TTACTTTCTT AAATAAAGTC TGTAAACTCA 10080

ATCGTTTGCT TGGTATTTTA ATTACCTTGT GTACTTTGGT CTGGGGAATG GTCATAGGTG 10140

TTGTCTATCT CTTACCTATT TTGATTAATC AGTTATCTAG TTTGATTATA TCTAGTCAAA 10200

CTATTTATAG TCGAGTACAA GACTTAATCA TAGACTTATC TAATTATCCT GCGCTCCAGA 10260

ATTTGGATGT AGAAGCTACA ATTCAGCAGT TAAACTTATC CTATGTTGAT ATTCTTCAAA 10320

ATATCCTAAA TAGCGTATCA AATAGTGTGG GGAGCGTCTT GTCAGCTCTT ATCAGTACTG 10380

TTTTGATTTT GATTATGACT CCAGTTTTTT TGGTTTATTT CTTATTAGAT GGACATAAAT 10440

TCTTGCCCAT GCTTGAAAGA ACGATTCTAA AGAGGGATCG CTTGCATATT GCAGGCTTAT 10500

TAAAGAATTT AAATGCGACG ATTGCTCGCT ATATTAGTGG AGTTTCGATT GACGCAATCA 10560

TTATAGGTTG TTTGGCTTAT ATTGGCTATA GTATTATTGG TTTAAAATAT GCTTTAGTTT 10620

TTGCCATTTT TTCTGGTGTA GCCAATTTAA TTCCTTATGT GGGGCCAAGT ATTGGTTTGA 10680

TTCCTATGAT CATCGCAAAT ATATTCACTG ATCCCCATAG ACTGCTGATT GCAGTGATTT 10740

ATATGCTTGT TGTTCAGCAG GTAGATGGCA ATATCTTATA TCCTCGAATC GTAGGAAGTG 10800

TTATGAAGGT TCATCCAATC ACGATTTTAG TTTTACTTTT GTTGTCAAGC AATATCTATG 10860

GTGTAGTTGG AATGATTGTC GCAGTGCCAA CCTATTCTAT CTTGAAAGAA ATTTCTAAGT 10920

TCTTATCCCA TTTGTATGAA AATCATAAAA TAATGAAAGA ACGAGAAAGA GAATTAGCTA 10980

AGTAAAAGTC AGGAGAACCC TGATTTTTCT TTACTGGAAG TGGCCTTTAG ATTAGAAGAC 11040

TGAAAATAAG TTAAAGTCTT AAACTAATTT TCACAGCTAA GAATAGTAGA AGTTAATCTG 11100

ATAAAAATCG AAAAAACCAG TGGAATTCTG TGTCAGGGTA AGTTCCACTG GTTTTCATAG 11160

TCTATTAAAG TTCGAATGAA ACCTATTTAT AGTAGATTGA AACTAGAATA GTACACCTCT 11220

AATTCTAAAA CATTGTTAGA AATCGATTTG ACTGTCCTGA TCTATTCGTT CTATTCTTAT 11280

TTCATTTTAC TATATTTTGG TGCAATAAGT GAAAAGTAGT CCGAATAATA TAAGGATTGA 11340

TTTTATAGTT TTTAAACTCA AATGAATTGA AATAAAGAGA GTACGAAAAT TCTCATCTGA 11400

AAGTATTTTA GAATAATTCT CTTCGTGAAT TTCTTCAAAA CAGATAGCTT CATCTTAGGT 11460

ATGTGATTTC TTTTTGCATT TTTGAGTTAG ATAAGGTATA ATGATTTTAT TGTCTTTTGG 11520

GGTCGTTACG GATTCGACAG GCATTATGAG GCATATTTTG CGACTCGTGT GGCGACGTAA 11580

ACGCTCAGTT AAATATAACT GCAAAAAATA ACACTTCTTA CGCTCTAGCT GCCTAAAAAC 11640

CAGCAGGCGT GACCCGATTT GGATTGCTCG TGTTCAATGA CAGGTCTTAT TATTAGCGAG 11700

ATACGATTAA GCCTTGTCTA GCGGTTTGAT AAGAGATTGA TAGACTCGCA GTTTCTAGAC 11760

TTGAGTTATG TGTCGAGGGG CTGTTAAAAT AATACATAAC CTATGGTTGT AGACAAATAT 11820

GTTGGCAGGT GTTTGGACGT GGGTTCGACT CCCACCGGCT CCATTATTCC TTTGCATTCT 11880

TTTGCATTCC TTGGTAAAAC GTTGTTAAAT CAACGTTTTT TATTTTTATC TTTGGTATTC 11940

CTTTGCATTC TTTTGCTAAA AAGGGAGTCA CAAACAGACC CTATTTTAAA AAAGGATAGA 12000

AAAAAGGATA CAACATTTGT CGCATCCTAA AAATAATCTT TTTTCGACGG AAGACATGGG 12060

ATTCGAACCC ACGCACGCTA TTACACGCCT ACCGCGTTTC CAACACGGCC TCTTAAGCCT 12120

CTTGAGTAAT CTTCCAATAC TTACTCAAAT AGTCTACCAT AAAGGCTCTT ATCTTGCAAT 12180

AAAAATTCTA GAAATAAGAA AAATGATAGA TTTTGAAAGA AAATGATAAA AAATGCTTGA 12240

CTTCGAAAGA AAGTATGATA GAATGAATAG TGTAAACGAT AACAGGAGGT GATTCAGTGT 12300

TAAAAACAGA ACGTAAACAA CTAATTTTAG AGGAGTTAAA TCAACATCAT GTAGTTTCTC 12360

TAGAAAAATT AGTTAGTTTG CTAGAAACGT CAGAATCAAC GGTTCGAAGA GACTTGGATG 12420

AGTTGGAAGC GGAAAACAAG CTTCGTCGTG TGCATGGTGG AGCAGAACTC CCCTACTCCT 12480

TACAGGAAGA AGAAACCATT CAAGAAAAAT CTGTCAAAAA CCTTCAAGAA AAGAAATTGC 12540

TGGCTCAGAA AGCAGCCTCT CTCATTAAAG AAAAAGATGT CATCTTTATC GATGCTGGAA 12600

CAACAACTGC TTTTTTGATT CATGAATTGG TCAATAAGAA TGTTACAGTT GTGACCAACT 12660

CCATTCACCA TGCCGCTCAG TTGGTTGAAA AGCAGAwTCC AACTGTCATG GTTGGAGGAA 12720

ACGTCAAGAC GGCGACAGAT GCTAGTATCG GGGGCGTTGC TCTTAACCAG ATTAACCAAT 12780

TGCACTTTGA CCGTGCCTTT ATCGGAATAA ATGGTGTTGA CGATGGCTAT TATACGACTC 12840

CTGATATGGA GGAGGGAGCT GTGAAAAGAG CTATTTTGGA GAATGCCAAG CAGACCTACG 12900

TCTTGGTGGA TTCGTCAAAA ATTGGACAAA CTTGCTTTGC CAAGGTAGCC CCACTCAAAC 12960

GCGCTATCGT TATCACTAGT CAAGGGCATG AGCTCTTGCA GGTTATTAAG GAGAAAACGG 13020

AGGTAATAGA AGTATGATTT ATACAGTCAC ACTCAATCCA TCCATTGACT ATATCGTTCG 13080

TTTGGACCAA GTCAAAGTTG GTAGTGTCAA TCGTATGGAC AGTGATGATA AGTTTGCTGG 13140

TGGGAAAGGA ATCAATGTCA GCCGTGTCTT GAAACGTTTG AATATACCAA ATACAGCGAC 13200

GGGATTTATC GGTGGCTTTA CTGGTAAATT TATCACAGAT ACTTTAGCAG AGGAAGAAAT 13260

CGAGACACGT TTTGTCCAGG TGGCAGAAGA TACTCGTATC AATGTTAAAA TCAAAGCAGA 13320

CCAAGAAACA GAAATCAACG GAACGGGTCC AACTGTTGAA TCGGTTCAGC TAGAAGAATT 13380

GAAAGCTATT TTATCTAGTC TGACAGCAGA AGATACAGTT GTCTTTGCAG GTTCAAGTGC 13440

TAAAAATCTA GGCAATGTTA TCTATAAGGA TTTGATTTCC TTGACGCGCC AGACTGGTGC 13500

GCAAGTGGTC TGTGACTTTG AAGGACAGAC CTTAATTGAT AGTTTGGACT ACCAGCCTCT 13560

TCTTGTAAAA CCAAACAATC ATGAACTTGG AGCGATTTTT GGGGTTAAAC TCGAAAGTTT 13620

AGATGAAATT GAGAAATACG CTCGTGAGTT ACTGGCTAAG GGTGCTCAAA ATGTTATTAT 13680

CTCTATGGCT GGTGATGGTG CCCTTCTTGT CACATCTGAG GGAGCTTACT TCGCTAAACC 13740

AATCAAAGGA ACAGTCAAAA ATTCAGTTGG AGCTGGTGAT TCTATGGTTG CTGGATTCAC 13800

AGGTGAATTT GTCAAATCAA AAGACGTAGT AGAAGCCTTC AAATGGGGAG TGGCTTGCGG 13860

AACGGCAACT ACCTTCTCAG ATGACTTGGC AACGGCGGAA TTTATTAAAG AAACATATGG 13920

AAAAGTTGAG GTAGAAAAAC GATGAAAATT CAAGACCTAT TGAGAAAAGA TGTCATGTTG 13980

CTAGATTTGC AGGCAACTGA AAAAACAGCT GTCATCGACG AGATGATTAA AAATTTGACA 14040

GACCACGGTT ATGTAACAGA TTTTGAAACA TTTAAAGAAG GAATTTTGGC GCGTGAAGCT 14100

TTGACTTCTA CTGGTTTGGG TGATGGAATC GCAATGCCTC ACAGCAAAAA CGCTGCTGTC 14160

AAAGAAGCGA CAGTTCTATT TGCTAAGTCA AATAAGGGTG TTGACTACGA GAGCTTGGAT 14220

GGACAAGCAA CTGACCTCTT CTTCATGATT GCAGCTCCAG AAGGTGCCAA TGATACTCAC 14280

TTGGCAGCCT TGGCAGAATT GTCTCAATAC TTGATGAAAG ACGGTTTTGC AGACAAACTT 14340

CGTCAAGCAA CATCTGCAGA CCAAGTTATC GAACTTTTTG ACCAAGCTTC AGAAAAAACT 14400

GAGGAACTTG TTCAAGCACC TGCTAATGAC TCTGGTGACT TTATCGTAGC TGTTACAGCT 14460

TGTACAACAG GTATTGCCCA CACTTACATG GCCCAAGAAG CCCTTCAAAA AGTAGCTGCT 14520

GAAATGGGGG TTGGTATCAA GGTCGAAACC AACGGTGCTA GCGGTGTTGG AAATCAACTA 14580

ACTGCAGAAG ATATCCGTAA GGCTAAAGCT ATTATCATTG CAGCAGACAA GGCCGTTGAA 14640

ATGGATCGAT TTGATGGAAA ACCATTGATC AATCGTCCAG TTGCTGACGG TATCCGTAAG 14700

ACAGAAGAGC TAATTAACTT GGCTCTTTCA GGAGATACTG AAGTCTACCG TGCCGCTAAT 14760

GGTGCCAAAG CTGCAACAGC CTCTAACGAA AAACAAAGCC TTGGTGGTGC CTTGTACAAA 14820

CACTTGATGA GTGGTGTATC TCAAATGTTA CCATTCGTTA TCGGTGGTGG TATCATGATT 14880

GCCCTTGCCT TCTTGATTGA CGGTGCTTTG GGTGTTCCAA ATGAAAACCT TGGCAATCTT 14940

GGTTCTTACC ATGAGTTAGC TTCTATGTTC ATGAAAATTG GTGGAGCTGC CTTTGGTTTG 15000

ATGCTTCCAG TCTTTGCGGG TTATGTTGCC TACTCTATTG CTGAAAAACC GGGTTTGGTA 15060

GCAGGTTTCG TGGCTGGTGC TATTGCCAAA GAAGGTTTTG CCTTTGGTAA AATTCCTTAT 15120

GCCGCAGGTG GTGAAGCAAC TTCAACTCTT GCAGGTGTCT CATCTGGTTT CCTAGGTGCC 15180

CTTGTTGGTG GATTTATCGC AGGTGCCTTG GTTCTTGCCA TCAAGAAATA CGTTAAAGTT 15240

CCTCGTTCAC TCGAAGGTGC TAAATCAATC CTTCTATTGC CACTTCTTGG AACAATCTTG 15300

ACAGGATTTG TTATGCTAGC TGTGAATATC CCAATGGCTG CAATCAACAC TGCTATGAAT 15360

GACTTCCTAG GCGGTCTTGG AGGAGGTTCA GCTGTCCTTC TTGGTATCGT CCTTGGTGGA 15420

ATGATGGCTG TTGACATGGG TGGACCAGTT AATAAAGCAG CTTATGTCTT TGGTACAGGT 15480

ACGCTTGCAG CAACTGTTTC TTCAGGTGGT TCTGTAGCCA TGGCAGCAGT TATGGCTGGA 15540

GGAATGGTGC CACCACTTGC AATCTTTGTC GCAACTCTTC TTTTCAAAGA TAAATTTACT 15600

AAGGAAGAAC GTAACTCTGG TTTGACAAAC ATCATCATGG GCTTGTCATT TATCACTGAG 15660

GGAGCGATTC CATTTGGTGC CGCTGACCCA GCTCGTGCGA TTCCAAGCTT CATCCTTGGT 15720

TCAGCAGTAG CAGGTGGACT CGTTGGTCTT ACTGGTATCA AACTCATGGC GCCACACGGA 15780

GGAATCTTCG TTATCGCCCT TACTTCAAAT GCTCTCCTTT ACCTCGTTTC TGTCTTGGTA 15840

GGAGCAATCG TAAGTGGTGT GGTTTATGGT TACCTACGCA AACCACAAGC ATAAAAAATA 15900

GAAAAATGAA AAGATTGGAC CGTTTGGTGC AGTCTTTTTC TCTTCCCGAA ATGCCTGTGA 15960

AATATGGTAT AATAGAAGAA TGGCAAACAA GAATACAAGT ACAACAAGAC GGAGACCGTC 16020

TAAAGCAGAA CTGGAAAGAA AAGAAGCGAT TCAACGAATG TTGATTTCGT TAGGAATTGC 16080

GATTTTATTG ATTTTCGCAG CCTTCAAATT AGGGGCTGCA GGTATAACCC TTTATAATTT 16140

AATTCGCTTG CTAGTGGGTA GCCTAGCTTA TCTGGCGATA TTCGGCCTAT TAATCTATCT 16200

CTTCTTTTTC AAGTGGATAC GAAAACAGGA AGGACTCTTA TCTGGCTTTT TCACCATATT 16260

TGCTGGCTTA CTCTTGATTT TTGAGGCCTA CTTGGTTTGG AAATATGGTT TGGACAAGTC 16320

CGTTCTAAAA GGGACCATGG CTCAGGTTGT GACAGATCTG ACTGGTTTTC GAACGACTAG 16380

CTTTGCTGGA GGGGGCTTGA TCGGGGTCGC TCTTTATATT CCAACAGCCT TTCTCTTTTC 16440

AAATATCGGA ACTTACTTTA TTGGTTCTAT CTTGATTTTA GTGGGTTCTC TCCTAGTCAG 16500

CCCTTGGTCT GTTTACGATA TTGCTGAATT TTTCAGTAGA GGCTTTGCCA AATGGTGGGA 16560

AGGGCACGAG CGTCGAAAAG AGGAACGCTT TGTCAAACAA GAAGAAAAAG CTCGCCAAAA 16620

GGCTGAGAAA GAGGCTAGAT TAGAACAAGA AGAGACTGAA AAAGCCTTAC TCGATTTGCC 16680

TCCTGTTGAT ATGGAAACGG GTGAAATTCT GACAGAGGAA GCTGTTCAAA ATCTTCCACC 16740

TATTCCAGAA GAAAAGTGGG TGGAACCAGA AATCATCCTG CCTCAAGCTG AACTTAAATT 16800

CCCTGAACAG GAAGATGACT CAGATGACGA AGATGTTCAG GTCGATTTTT CAGCCAAAGA 16860

AGCCCTTGAA TACAAACTTC CAAGCTTACA ACTCTTTGCA CCAGATAAAC CAAAAGATCA 16920

GTCTAAAGAG AAGAAAATTG TCAGAGAAAA TATCAAAATC TTAGAAGCAA CCTTTGCTAG 16980

CTTTGGTATT AAGGTAACAG TTGAACGGGC CGAAATTGGG CCATCAGTGA CCAAGTATGA 17040

AGTCAAGCCG GCTGTTGGTG TAAGGGTCAA CCGCATTTCC AATCTATCAG ATGACCTCGC 17100

TCTAGCCTTG GCTGCCAAAG ATGTCCGGAT TGAAGCACCA ATCCCTGGGA AATCCCTAAT 17160

CGGAATTGAA GTGCCCAACT CCGATATTGC CACTGTATCT TTCCGAGAAC TATGGGAACA 17220

ATCGCAAACG AAAGCAGAAA ATTTCTTGGA AATTCCTTTA GGGAAGGCTG TTAATGGAAC 17280

CGCAAGAGCT TTTGACCTTT CTAAAATGCC CCACTTGCTA GTTGCAGGTT CAACGGGTTC 17340

AGGGAAGTCA GTAGCAGTTA ACGGCATTAT TGCTAGCATT CTCATGAAGG CGAGACCAGA 17400

TCAAGTTAAA TTTATGATGG TCGATCCCAA GATGGTTGAG TTATCTGTTT ACAATGATAT 17460

TCCCCACCTC TTGATTCCAG TCGTGACCAA TCCACGCAAA GCCAGCAAGG CTCTGCAAAA 17520

GGTTGTGGAT GAAATGGAAA ACCGTTATGA ACTCTTTGCC AAGGTGGGAG TTCGGAATAT 17580

TGCAGGTTTT AATGCCAAGG TAGAAGAGTT CAATTCCCAG TCTGAGTACA AGCAAATTCC 17640

GCTACCATTC ATTGTCGTGA TTGTGGATGA GTTGGCTGAC CTCATGATGG TGGCCAGCAA 17700

GGAAGTGGAA GATGCTATCA TCCGTCTTGG GCAGAAGGCG CGTGCTGCAG GTATCCACAT 17760

GATTCTTGCA ACTCAGCGTC CATCTGTTGA TGTCATCTCT GGTTTGATTA AGGCCAATGT 17820

TCCATCTCGT GTAGCATTTG CGGTTTCATC AGGAACAGAC TCCCGTACGA TTTTGGATGA 17880

AAATGGAGCA GAAAAACTTC TTGGTCGAGG AGACATGCTC TTTAAACCGA TTGATGAAAA 17940

TCATCCAGTT CGTCTCCAAG GCTCCTTTAT CTCGGATGAC GATGTTGAGC GCATTGTGAA 18000

CTTCATCAAG ACTCAGGCAG ATGCAGACTA CGATGAGAGT TTTGATCCAG GTGAGGTTTC 18060

TGAAAATGAA GGAGAATTTT CGGATGGAGA TGCTGGTGGT GATCCGCTTT TTGAAGAAGC 18120

TAAGTCTTTG GTTATCGAAA CACAGAAAGC CAGTGCGTCT ATGATTCAGC GTCGTTTATC 18180

AGTTGGATTT AACCGTGCGA CCCGTCTCAT GGAAGAACTG GAGATAGCAG GTGTCATCGG 18240

TCCAGCTGAA GGTACCAAAC CTCGAAAAGT GTTACAACAA TAAAAAAATA GCTTCTTTCC 18300

AAGTTTGGAG GGAAGCTATT TTAGTGGCTA TTGATTGCTT TTATTTTCTG AAGTTGGCGC 18360

ATTGGACTGT TTTTCGTTTT CAGTAGCAGG TTTACTTGAA GCAGGAGTAG AAGAGTCCTG 18420

AGTTGCTGTT TTCTGATCTT CTTTTTTCTC TTCCTTGACG CTAGATTTTG GTGTTTCCTC 18480

TTGCTGTGTT TTTTCTTGAC TAGTGTTAGT CTCTTTAGTT GGACTGGTGT TTTCCTTAGG 18540

GGATTCCTTT TGGATTTCTT TGACAATGGT TGTCGTCTGG CTTGTCGTAG GTTCTTTTTT 18600

AATATTTTTG TTATTATCCA AGGCGTT 18627

(2) INFORMATION FOR SEQ ID NO: 114:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2560 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

TAAAATACGT TACCTTGCTT CTGCACGTTC AGCAGGTAAG TCATTGAAAT TTAAAGATCA 60

AGATATTACA ATTGAAGAAA CGACTGAAAC AGCTTTTGAA GGAGTTGATA TTGCTCTCTT 120

TTCAGCAGGT AGTTCTACAT CAGCTAAGTA TGCACCATAC GCAGTAAAAG CTGGCGTGGT 180

AGTAGTAGAT AATACATCTT ATTTCCGTCA AAATCCAGAT GTTCCTTTGG TTGTTCCAGA 240

GGTCAATGCT CATGCACTTG ATGCTCACAA CGGAATCATT GCCTGCCCTA ATTGTTCAAC 300

AATTCAAATG ATGGTGGCTC TTGAGCCGGT TCGCCAAAAA TGGGGCTTGG ACCGTATCAT 360

TGTTTCAACT TATCAAGCCG TTTCAGGTGC TGGTATGGGA GCAATTCTTG AGACACAACG 420

TGAACTTCGT GAAGTCTTGA ATGATGGTGT GAAACCACGT GATTTGCATG CGGAAATCTT 480

GCCTTCAGGT GGTGACAAGA AACATTATCC TATCGCCTTT AACGCTCTTC CACAAATTGA 540

TGTTTTCACT GATAATGATT ACACGTACGA AGAGATGAAG ATGACCAAGG AAACTAAGAA 600

AATTATGGAA GATGATAGCA TTGCAGTATC TGCAACATGT GTGCGTATTC CAGTCTTGTC 660

AGCTCACTCT GAGTCTGTTT ATATCGAAAC AAAAGAAGTG GCTCCAATCG AAGAAGTAAA 720

AGCAGCTATC GCAGCCTTCC CAGGTGCTGT TCTTGAAGAT GATGTAGCTC ATCAAATCTA 780

TCCTCAAGCT ATCAATGCAG TTGGTTCGCG TGATACCTTT GTTGGTCGTA TCCGTAAAGA 840

CTTGGATGCA GAAAAAGGAA TTCACATGTG GGTTGTTTCA GATAACCTTC TCAAAGGTGC 900

TGCTTGGAAC TCAGTTCAGA TTGCTGAAAC TCTTCATGAA CGTGGATTGG TTCGTCCAAC 960

AGCCGAATTG AAATTTGAAT TAAAATAGTC ATATCGTTTA GGAGTTCAGA TGAACTCCTT 1020

CTTTGAAATA GAGAGGTGTT TTCGTGTCTT ATCAAGATTT AAAAAAATGT AAAATCATTA 1080

CAGCCTTTAT TACCCCCTTC CATGAGGATG GTTCCATTAA CTTTGATGCT ATTCCAGCCT 1140

TGATTGAGCA TTTATTGGCC CATCATACGG ATGGAATTCT TCTCGCAGGA ACGACTGCTG 1200

AGAGTCCAAC TTTGACCCAC GATGAGGAGT TGGAGTTGTT TGCGGCTGTA CAAAAGGTTG 1260

TCAATGGACG CGTTCCTTTG ATTGCGGGTG TAGGTACTAA TGATACGCGT GACTCTATTG 1320

AGTTTGTCAA AGAAGTAGCG GAATTTGGTG GTTTCGCAGC TGGGCTTGCT ATTGTTCCTT 1380

ACTACAACAA ACCTTCTCAA GAAGGGATGT ATCAGCACTT TAAGACTATT GCAGATGCTT 1440

CTGACCTACC AATTATTATC TATAACATTC CAGGGCGTGT AGTTGTCGAA TTGACTCCAG 1500

AAACCATGCT TCGCTTGGCT GACCATCCAA ATATTATCGG TGTCAAAGAA TGTACTAGCT 1560

TGGCTAATAT GGCTTACTTG ATTGAGCACA AGCCTGAAGA GTTCTTGATT TATACAGGTG 1620

AGGATGGAGA TGCTTTCCAT GCCATGAACC TTGGGGCGGA TGGGGTTATT TCTGTTGCCT 1680

CTCATACAAA TGGGGATGAA ATGCACGAGA TGTTTACTGC GATTGCAGAA AGCGATATGA 1740

AGAAAGCCGC AGCAATTCAG CGTAAATTCA TTCCTAAGGT TAATGCTCTC TTCTCTTATC 1800

CAAGTCCTGC TCCAGTTAAG GCAATTCTTA ACTATATGGG ATTTGAAGCT GGACCCACTC 1860

GTCTACCTCT TGTTCCAGCA CCAGAAGAAG ATGCCAAACG CATTATCAAG GTTGTCGTAG 1920

ATGGCGACTA CGAAGCAACT AAGGCAACTG TAACAGGGGT CTTAAGACCA GATTACTAAT 1980

AAAGACAATA AAATCCGGCT CTTTGTCAAC TGTAGTGGGT TGAAGTCAGC TAAGCTCGAG 2040

AAAGGACAAA TTTTGTCCTT TCTTTTTTGA TATTCAGAGC GATAAAAATC CGTTTTTTGA 2100

AGTTTTCAAA GTTCCGAAAA CCAAAGGCAT TGCGCTTGAT AAGTTTGATG AGATTATTGG 2160

TCGCTTCCAA TTTGGCGTTT GAATAGGGTA GTTGAAGGGT GTTGACGATT TTCTTTTTGT 2220

CCTTTAGAAA GGTTTTAAAG ACAGTCTGAA AAATAGGATG AACCTGCTTC AGATTGTCCT 2280

CAATGAGTCC GAAAAATTTC TCCGGTTCCT TATTCTGAAA GTGAAACAGC AAGAGTTGAT 2340

AGAGCTGATA GTGATGTTTC AAGTTTTGTG AATAGCTCAA AAGCTTGTTT AAAATCTCTT 2400

TATTGGTTAA GTGCATACGA AAAGTAGGAC GATAAAATCG CTTATCACTC AGTTTACGGC 2460

TATCCTGTTG AATGAGTTTC CAGTAGCGCT TGATAGCCTT GTATTCGGGA TTTTCGATGA 2520

AACTGATTCA TGATTTGGAC ACGCACACGA CTCATAGCAC 2560

(2) INFORMATION FOR SEQ ID NO: 115:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11303 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

TATTGGATTT CCCTTGCAAT CAGTTTATGG GACAAGCACC CGGCAGCGCA GAGGAAATCA 60

ACGCCTTCTG TAGCCTACAT TTTCAAACCA CCTTCCCACG TTTTGCCAAG ATTAAGGTCA 120

ACGGTAAGGA AGCAGACCCT CTCTATGTCT GGTTACAAGA CCAGAAATCC GGCCCACTAG 180

GAAAACGAGT CGAATGGAAT TTCGCTAAGT TTCTCATCGG TCGAGATGGG CAAGTCTTTG 240

AACGCTTTTC TTCAAAAACA GACCCAAAAC AAATTGAAGA GGCGATACAA ACTCTACTAT 300

AATTCACAAT CTCACTATGA TTAGGTTTCC TTTAACCTGA TGAATAGTGA GATTTTTTGA 360

TGGGCTTTGA CTTAAATAGA AAAACACCCC ATGATATGAA ACATGAAGTG TTGTAAAGTC 420

TATGTTGTAG GTGCTTATTT CACAATTTCA ATGTGACCAG TGATAACGAA TACCATACAG 480

AATCTTCATA TACACTAAAC AAATGACTTT CTAATTATTT CAATTAGTTT TGGCTAGTAA 540

ATATCATTTC CAACAAACGC CCTCTCAATT CCTTATCCTG ATGATGCAAG ATATTCATTA 600

AGTCATGAGA GTTTTTCGCA TTGATGAATT GATTTAACAA TCTATCTTTT AATTCATATG 660

GAAGAGAAGC TGTCTTTAGT AGTCTAAAAA CTTCGTCATT TAAAGATGTC CTTTTATTAT 720

CTTTCCATTC AAATTTAGCT GTATCATTCT TATTTGGCAA TTCAATTATA GACACATTCG 780

TTCCTTTAAA ATGAATTCTA TGTTTTCTAT TGCTTGGAAC GATACTAGAA TCTCCTTGTA 840

ATGCTAACTC TACCATTCCC ATTTCCCAAT CGATTGATAA TCTTGTTTTA TATCTTTGAC 900

CATTTTGATC TTCAAGCATT TCAAAAGAAT GTTGTTTTCC TGGGAATACA TACCAATCTA 960

CAACTTCAGG TAAATCAACA CCCATACCTA TCTCAGAACC AACCAAGGGA ATGATTGCAC 1020

CACTTTTTGC AAACACAGGC GTAGTCGAGA TGTCCCTATA AACACTTAAC TTCACACCAC 1080

CTGTGTATTT TTTCTCTGAA AAGAAGTCAT ACCATTCACC TTCAGGGAAC CATACATCTA 1140

CTTTTGCAGA TTGGAATGTC AAATCCATCT TTTCTACAAT GGGAGCCACC ATCAGTTCTG 1200

TTCCAAAAAA GTATTGGTTT GGAACATTAT AGCTCTCATC ATTCTCTGGA TAGAAATAAT 1260

AGATTGGACT GATTAATGGG GCACCTTCCT CATGTGTCTG TACATTCATG GTATATAGAT 1320

AGGGAATCAT CTGATGTCTC AAACGAAGGT ATTTCTTCAT AATCTTAGAT GTTGTTTCTG 1380

AAAAAAACCA AGGTTCTTTA CTATTAAAAG GACTTCTAGA ACTATGTAAT CGAGTAATCG 1440

GACTAAAAAC ACCAAACTGT AGCCATCTAG TTTGTAGCTC TTCGTCATAA TCCCCCAACA 1500

TATGTCCACC GATATCATGA CTCCACCAAC TATAACCGAT ATTAGATGCT GTCGCTGTAA 1560

AATAGGGTTG AAATCTTAAG GAATTCCAAC TAATAATAGT ATCCCCTGAA AAACCAACAG 1620

GGTAGCGGTG ACTACCAGGA CCTGCATATC TTGATAAAAT CAAACCACCT TCTGCATTTT 1680

TACAACTATC CTGATAGTGA TAATGGTTTA AAAGCCAAAG TGGATCTAGC ATACCTTGTG 1740

TCCCTTGTTG CCAGTCAATC CACCAAAAAT CTACTCCCTG CTTTTCTAGT TCATAATGAA 1800

CATCTTTAAA GTAGGCTTCC CTAAAAGAGG GATTAAAAAA ATCAAAAATA GCAGGTTCTT 1860

CTAGTTCTAC ATTTAACCCC AACCGTTTTG CGATTTGAGG ATAAGCTTCT TCATAAGCCC 1920

GTATCCCATC AGCAGGATGG ACATTTAAGG AGAGTTTTAG CTTTCTATCA TGAAGTTGTT 1980

GCAATAACTG TTCTGGATTT GGTATTAAGT TTCTATTCCA ACTATATCCT GTCCAGCCAC 2040

TTCCAAAGCG AGCTGGAATG TCAGTTATAT GCCAATCCAT ATCTAACACA CCGATAGATA 2100

ATGGAATTTT CTCTGTTTCA AATCTGTCTA TTAAATCCAA GTATTCATCC GACGTATAAG 2160

GCCAATATCT ACTCCACCAA TTGCCTAAAG CATATCTTGG CAACAAGGGT GTTGAACCAG 2220

TCAAATGGTA AAAATCTCTG ATTGCTCCTC TATAATCATG CCCATAGGCA AAGAAATACA 2280

GGTCAATTTG ATTTTCTCTC TCAATATAAC CAGATTGTTC ATCCCAAATA AATCCTTGAG 2340

AATCATCCAA TAAGGCTATA CCATTTCGGC TAATAATTCC ATCTTCTAAC GAGATTGCTC 2400

CATCTGCCTT ATCCAGAGTC CGAGCTGTTC CTTTTAACGT TTCAATAGAT TCACCAAAAT 2460

ACCAGCGACT ACCATATACG GCAAAATTTC CTTTTAATTC TATAAATAAA TTTTCGGCGT 2520

TAAATTCTCC TTTATTAAAG TGCAGATGAA AATAGTCCGT CATAATATCT AGTACGTTTG 2580

ATGTCTCGAT ATAATCTAAC GAAATTTGGC CAAAATCTCT ATTATAGATA AGTTGTGTCG 2640

TTCTATCCTC AAAACTTCCA GTTTGAGAGT ATTCTAACCT TACTAGCTTG TCTGTTAATA 2700

CAGAGATTCG ATAAAACTCT CCCTTAAAAA TTTTCAATTT GTTTTCCTCC TTTTATGGTA 2760

GCATAAAAAC AGAACGCACC ATTTTTGATG CGTTTTTCAT TATTCTGAAT GCAATGTTCT 2820

ATCTGTTATA TCTATGACAA ATAATAGTCA ATTGAAAAAA TGCAGTGGAC AAAATATCTT 2880

TTAACAAACC AAGAGTTTAT TAAAGAGTTA TCACTTTTCA ACTTTTCTAA GCTTATGCAG 2940

TTGTGAAACA AACTACTTTT AAACTATTAA CTAAGATAGG ATTGATAAAT AATTTCAAAC 3000

TCTTACTAGC AATCATACGA TATTCAAGCT CACGTGCTTT TTTCCTTCCT GCTTATTTCT 3060

TAGAACTGAA GAACCCGGAT CGGTATATAA ATTATCCGGA TCAACATAGT CATAAGATTC 3120

ATAACAGTTG CGCTTCATTA AGTCATCCCC AGAGCAAGAG CTTCATCTCG TAATTTTTCA 3180

ACATCACTAA CCGTAGGTCG CCATCCTTCA ATCATATTTG TACTTAAAGC ATACCAAACA 3240

CTCTTAAAAA CGGATCGGTT TTCAAAAGCT ATTCCCATGA TTGTCATCTT TTCTTTATCT 3300

ATATCTAAGG ACATATGCTA CCTCCTTTAG ATACATTATA CCATGTTTCT CTGTAGCTTT 3360

TAAAAATTTT ATTTTGTTTG TCATATCTAA GTTTTCAGCA CGCTTATCCT ATTTTATAAG 3420

CCTCAAACCC AAATATAAAA CGCATTCTTT TTGCTTTTTT ACTATTGTAT CGTATTCTAC 3480

GATAACATAC TTTACTTTAT TGTTTTTTTA AATAACAGCA GTTCCCTGTT TATCAACTAT 3540

TCGAACTACT TTCTATTTTG CTTCATACCC TACATAGCGA AAAAATATGA AAAAGCAGAG 3600

AAGAATATCT TAAAAAGACC TCTTCACTGC TAATATTAAC ACTCATTATT TAAACTATAT 3660

GGATTCTATC ATCGAGTATA CTTTTTTACT TATTAGATAC CTTGCTCTTC TTTCACCAAT 3720

TTTTGATCAT ATACACGGAT GAATGGAAGA TAGACTAGGA ATGCTGCAAA TGCACATACT 3780

AGAGCAACTA ATACAGCTCG AAGATCTGCT GTCCCTAAGA AAGCTCCAAT CCCTACTGGA 3840

GTTGGCCATG GAACCTGTGC GATAATTGGC TTAATAAAGT TTAGAGAATT CGCTACGTAA 3900

TAAATAGTAG CAGTAACCAT TGGTGCTAAA ATAAATGGTA TAGCCAAGGC TGGATTATAG 3960

ATAATAGGTA ATCCAAAAAT TAATGGTTCA TTAATATTAA ATAAGGCTGG AACTACAGAT 4020

GCTCGTCCTA TTGCTTTAAG CTGTTCAGAT TTAGAGGCAA AAGCAATATA TAAACATAGT 4080

CCTAAAGTTG CACCAGAACC ACCTGCAATT ACAAACATAT TAGAAAATTC ACCTGCAACA 4140

GCGAAGTGCC CGCCAGCAGC ATTTTCAGCC ATGTTAGCAA GAGCAATTGG ACTAACAAAT 4200

GCAAAAACAA TGTTCGCACC GTGGATACCT ACAATCCAAA GTAGTTGAGT CAATAGATAA 4260

ATAATCATTA AACCAATCCA CGAATTAGTC AGATTGGATA CAAAACCAAA TGGAATTGCA 4320

ATGACTTTAA AAATATCTGT TCCCATTGCT ACAAGAAGAC CGTTGATAAA GATAACAACA 4380

AATGCAACAA CAAATCCCGG AACCAAAGCG GTAAATCCAC GAGAAACTCC TTCTGGAACA 4440

GCTTCAGGCA TTTTAATAAC CCAATTATGT TTAACACACA TACGATAAAT AAGAACAGTC 4500

ACAATTGCCA TAATGATTGC GGTAAAAATC CCTGTTGTCC CAAAACGTGC GACTACATTT 4560

CCCATTGCCC ATCCATCTGC AATTACTGCA CCTTCTTTTA GACTTGTCAC AGTCTTCATC 4620

ATTCCACCAT CAAAAATGAT TTGCGGTACT GTCATGACAA AAGCCATCAA GGCAAGCAAG 4680

GCACCATTAA GAGGATTCAT ATTGAGTTCT TCTTCCTCTG CATAAATTTT TGTCAATTCA 4740

TATGCAAGTG ATAGAACGAA ATAAAGAGAT AGAGAACCCA TAGTCGCATA GTTTGCAACC 4800

ATGTAAAGTG ATGTGAATTT ATCAAATGAA GCAGAGAAAA TATCTGCCAC AATTGGCCAA 4860

AATGAGAAAG CTTGTGGCAA AATACTGAAT ACCAAAAACA TTGATCCTAC AATAGTAAAT 4920

GGTACAGCAG CCATACCTGC AGCCGTGATA GCACGTACTA CTTTAAACTG AGCAAGTTTG 4980

CCCATTGGTC CCATAACATG GTTTTCAAGA AAACCAAACA ACCCGTTTTG TTGATCCATA 5040

AATAGACCTC CTTAATAAAA CATAATAATT TTTACTTTCT AAAGACTAGT TTCAAATACA 5100

AATTATACTA GATCAGGATT ATAAACTAAG TGAGTTCTTT TCCAATTGGA CAAATTGTTG 5160

ATAAGCCTTA TCTGTTCGTT TATAAATTTT TTTAATTCTT CTAATGTCTA ACAAACTCAG 5220

AACTAAACCT AATAGAAGAA CTACAAAAAC AAATAAACGT GCTACTTGGT TATTTTCAAA 5280

AATCGGAAAA AGATTCTTAA ACCAACTTGT CCAAGTTAAA ACAAGTAATC CTATTGAAAT 5340

AAGCATTTGT ATTCTAACAA ACATTAGTGT TATTCCCAAC TTTTCTTTCC TATTTCCATA 5400

AAGTTTAAAT TGTTCAACAG TTGCTAAAAT AGAAAATACT ATGAGCATAA TGGGGAAAAT 5460

AATAATAGGC GAGGGACTAA TAAACTGACT CAAAAGCCAA TAAATATTCC CAAAAAAGAA 5520

GAGTGCTATT GAATAACGTA GAAGAAGATA TCGATTGAAA AAAGTATTAG TTAGAGCCAT 5580

CTCTCGACGT TGTTGTTCAA TCTTTTGTCG TTCTTTTTTA TCCATATCAT TTCCTCCTTA 5640

TATAACAACA CATATTTAGT TAACTTTCTT ATAAAGAGCT AACATTTCCT TTGCTACTTC 5700

TAATAATGTC ATAGTGGTCA TTAAATGATC TTGAGCATGT ACCATGATAA TTTCAATTTT 5760

AATTTCCACT CCACTTGCGT ATTCTTGCAA GAGTTTGGTT TGTGCATGAT GCGCTTCAAG 5820

AATTATCTCA TTTGATTGAT TTAATTTACT TTCTGCATCA TCAAAACTAC CTTCTCTCAT 5880

TTTTGCAAAT GCTTCATGTA TTTCTGACCT TGCATTTCCC GAATGCAGGA TAATTTCAAA 5940

TGCTGCAACC TGCAGTTCCT CTTGATTCAT ATAAACCTCC TATTTTATCT TCTCAAATAT 6000

GTTAATAAAA TCTTCAAAGT TATTGCAAGA TATTAGCTGA TTTTGCAATT CATCATTCTC 6060

TGTCAGAGAG ACTATCTTTT TAGTCACAGT TGCCAAACCT TCGTTCCCAT ATATTGATGG 6120

AGATAGAAGA AATACTAGCT GGACATGTGA ACTTTGATTA TCCCAGAGTA ACGAATCTTT 6180

ACAAATTGCA ACCGAAACCT TTCCCTCTGT ACCAAAGGGC TGAATAGGAT GCGGAACTGC 6240

AATTTTTTCA GAAAAAACAA CTGAACTTAA TTCTTCGCGC TGTTTAATTC CATAAAGTAA 6300

AGATTGTTCA AACTCATTTG ATTCACCAAC AGATAAACTC TCAACCATCT TTTCAAGTAA 6360

ATTTACCTTG TCTGATTCAG TACATATTAA AAAGTTTTCT TTACTAAAAT ACTGTCTAAA 6420

GCCGTTGTTT TCAAATTTGT TAATCTTTGA TGATTGTACA TAACTAGAAA CTTGCATCTA 6480

ATCCATAGCT TTTCTAATCA TTTCCATCTC ATCACTCTTA AGAAACACAC TAACTTTAAA 6540

AACTGGGATT TGAAAATATA GATTTGATAA ATCAATAGCT GACACTATAA AATCTATTCC 6600

TTTAAGTTTT TCTTGATTCA ATTCATAGTA GCCTATTACA TCAACAACTT CTACTCGCTT 6660

CCCAAACTCC GTTTCCAAAC GATTTCTTAA CATTTGGGCT GCACCAAATC CTGTTGCACA 6720

AATAGCAAGA ATATTAAACT TAGTACTCTC TTTGCTACGT TCCATAGCAG CTAAAAAGTG 6780

AAGACTTACA TATGCTACTT CATCATCTGA TATTGTCCAC TCCAAGAACT TGTCCATATT 6840

TGCAAGAATT TCTCTAGTCA TAAAGAATAT ATCACTATAA TTCTGTTTAA TTTCATCTAC 6900

CAAAGGGTTA TTTAAGGTAA TCCGGCTTTC TAAACGTACT TGTAGTGTCA TTAGATGAGT 6960

TATCAATCCT TCAATTAGTT GGAAATCTGA AGAAAAGTTA TACATATCAT CTAATCCTAA 7020

ATTCTGAAAT GTTTTAAATA AAGATTTTTT TAAAACTTCT TCAGAAATAT TCTTCTGATT 7080

TTTTTGACAT TGTTGACTCT TAGCTAACAA ATGCAAAGTA ATGTAGTCTA TTTCCTGAAC 7140

TGGAAATTCC TGATTTGTTA CTTCTCTTAC TTTAGAAAGA ATTCTTTGGG CAACCTTTCT 7200

CTCTATTGCA TCATCAGTCA TCTGACAGTC TATATTTTTT ATTTCAAATC CGGATTTTAA 7260

ACGAATCACA GACAATGCTA TGTGAACTAC TAAATTCTGT AGTACAAAAT CAGATAGTTT 7320

TAGGTTGGCC TCTTGGCATT CATCCAAAAC AATTCTAGCA AATTCTTCTA ATGGAACAGT 7380

TTGATCAAAA AAGTTAAATT TTACATAGCA ATGTATTGTT TTAAAAAATT GATTCTCTAG 7440

GAAATAATTT ATGATAAAAC GTCGTTTATC ACGTTCCTCG CCTGAGACAT AAACTCCTTT 7500

ATTCGCCCTA CTCTCAATGG ACAAATTATA CTCTGATAAC ATCACTCGTA TCTTTCTGAA 7560

ATCATGAGAT AATGTTGAAC GACTAACGTA AAGTTCATCA GCTAAATCAT CAAAAAGAAC 7620

TGGAACTTGC TCAAATAATA ATTTATTTAA GATAAATACT AAACGATCAT CACCTTTTGA 7680

AACCGCAGTT TTCGTATAGT CTTCTTCCAG TTCATAAGTT TGTCTAAACT CCTGGTAAGC 7740

GCCTTGATTC TCAAAAAATA TTTGATACCC TTGACCTTGT TTTGAAATCA ACCGGACTCC 7800

TTGAATAATC ATTGTCTTCT CAATTAATTT CAGTACATTA CGGACAGTTC TATCTGAACA 7860

GGATAAATAT TCTGCCAGTT CTTTGCTTGT AACAAAACGT TCCTTATTTT TTATTAAAAA 7920

TTGAAGGATA TCTTTCTCTT TAATGTTTAA CACATTCATT CCCTCCTAAA ACGTATGTTT 7980

TCATATATTG AAGCATATTA TACACTTAAA TCAGTTTATA TCAAACTCAA AACAATTTAT 8040

CTTAACCTAA ATATTTATTG ACATTTCATG TGTTCATCAA ATATTCTCAA GAATCAAATT 8100

AGCCATTTTT TCAATTCCCA TTGGAATAGG AATATAGGCT TGAGGAGGTA TTTGTACAAC 8160

TGGTTTTCCT GCTTTAGAAC CAGCCTCTTC AAATTGCTTA AAGTACATTT TTGTTTGAGG 8220

ACTGACAAGA TACAAATCAA AAGCTGCTGC TGCGATAGCT TTCCCTCCTT CAGTAGCACT 8280

AATAGCATCA ACTACAATAT CTTTCCCTTT TCCTTTTAGA AACTCTGTTG TTTTCTGTGC 8340

CATAAGTGAT GAAGACATTC CTGCTGCACA AATAATTAAA GCTTTTGCCA TAATATTTTC 8400

TCCTTTTCTT AAATCCAATC AAAGCTGTGC TAAGTTGGCT TATTTGTTAT CTATTTTTAT 8460

TATAAAATAA AGCGTTTCCA ATGACAATTC CCTCATTTTC CTAAATGATA TGGAAAAAAA 8520

TTATTTATAC TTCAATTTAT AAAATAAAAT TATTCCTGAG AGTAGAAATG AAACACTATT 8580

TGCTAAAATC AAAGGCAAGT CTCCTATACG AATACCATGA GCAAGCCACA ATGCAATACC 8640

AATAACTTGC ATAACATACA TACCTAGAGC AATAGATCCT GTGTCCTTTG TCTTAACTAC 8700

ACGAAAAACT TGTGGTAAAA ATGCAAATGT TGTTAAAATT GCTGCAATAC TTCCAATCAT 8760

ATGTCACCTC AATATGCTAA ACAAACTGAG AATAATCTCA GTTTGTTTAT ACTATTCTAC 8820

TGATTCACCG TTAGATGAAA TAACTTCCTT ATACCAGCCA AAAGATTTTT TCGGGGAACG 8880

ATTATAACTT CCCTTCCCAT TATCATCTTT ATCTACATAA ATAAAGCCAT AACGTTTCCG 8940

CATTTCACCG GTACCAGCTG AAACCAAATC AATACATCCC CATGGAGTAT AACCCATTAA 9000

ATCAACACCA TCTTCAACTA CAGCCTTTTT CATTTCACGA ATATGGGCAC CTAGATATTC 9060

AATTCTATAA TCATCATGTA CCATACCATC TGCTGCAACT TGATCTATAG CTCCAAAACC 9120

ATTTTCAACA ATAAAGAGTG GTAAGTGATA GTGGTCTGTA AACCAATTTA ACGCATAACG 9180

CAAACCTTCT GGATCAATTT GCCACTCCCA TTCAGAAGCC TTAACATAAT TATTTTTCAC 9240

TAAATCTTCT GTTTCAAGAT AATCAAAATA AGGATTATTT TCACGATGAG AGTCGATAGC 9300

AAAGGACATA TAGTAACTGA AACCAATGTA ATCTACAGTC CCACCAAGTA AATCTTCTTT 9360

ATCCTGGGCA GTAAAATCAA CTGAAATACC TTTTCGTTCC CAATACTTGA AAATATGCTC 9420

AGGATATTTA CCTAAAACAT GCACATCAGC AAAATAATAA CGCTTCTGCA TAGCTTTCAT 9480

TGCCATTAAG ATATCCTTAG GATTGCAAGT AACTGGATAA ATTGGACACA TCGCAATCAT 9540

ACAACCTATT TGAAAATCTG GATTAATCTC ATGACCAATT TTTACAGCTC GTGCAGAAGC 9600

AACTAATTCG TAATGTGCTG CTTGATACAT AATTGCTTCT CTATTATCAC CTTCCTCATA 9660

TACAATACCT GAGTTAGTAA ATGGTGCAAA ATCTTCCTGA TAATTCGCTT GATTATTGAT 9720

TTCATTGAAA GTCATCCAAT ATTTAACCTT ATCTTTGTAA CGTTTAAATA CGACTTCTGC 9780

AAAACGAGCA AAGAAATCAA TCAATTTCCT ATTTTTCCAA CCACCATATT CGGTCACTAA 9840

GTGATAAGGC ATTTCAAAAT GAGATAGAGT GATGACAGGT TCAATACCAT TCTTTAAGCA 9900

TTCATCAAAA AGATTATCAT AAAACTGTAA TCCTTCTTCA TTCGGCTCTA ACTCATCACC 9960

TTTTGGAAAG ATACGTGTCC ATGCAATAGA GGTACGGAAG CACTTGAATC CCATTTCAGC 10020

AAAAAGTGCT ATATCTTCTT TATAACGGTG ATAAAAATCT ATCGCCTCAT GATTTGGATA 10080

ATATTTACCC TCTAAAACTC CCAAAGTAAT TTCACGAGCT ACTCCATGAC GACCAGCAGT 10140

CATAACATCA GCAACACTAA TTCCCTTGCC ACCTTCTTGC CATCCACCTT CAAGTTGATG 10200

AGCAGCAACA GCACCACCCC ATAAAAATCC ATCTTTAAAA GTAGTCATCT TTTTTCCTCC 10260

TGACTTTGAT ACTCTTATTA TAAACCTTAA ACCAAAAGAT GAAAACGCAT TCTTTTTCCT 10320

TATTGTTAAG GAAAGAAGTA ATTTTTAATG GAAATAGAAC AATATCTTCT TGTATTCTCG 10380

TAATGATATC TTTACGATTT TCAATACTTT CAAACTACAA AAACTCTCAC AATAATTCTA 10440

ATTCCCTGTG TCTATAAACG ACTTATCGCT TTCTGGCATC CCAGAATCAT CTTCTATATA 10500

ACGTTCAACT TGCATCTGCA AGTGATATTT TTTTCTTAAA TCTAAGATTT TCTGCATTGT 10560

CTTTGATTGA TAATGTTTAT CTAAAGTTTC TTGATTTATC CACTGATCAA TAAGGAGAAT 10620

AGTTCCCTCT TTTTCAATTG GTAAAAAATA TTCGTATTTC AAGTTACCTT TTTGATTTCT 10680

AATTTCTTTA ACAAGGCCAC TATCAAGCAT TTCTCTTGCA AACTTTATTG CACTATCTCC 10740

ATCACCTTTA TAATATACAT GAATAGTCAA TGTCATCTTA TATCCTCCAA AATCATCCTT 10800

CAATTTTAAA AAAACAAGTT TAGATGAGGA TCTAAACTTG TTTTTTATGA ACTAATTATC 10860

TAACGTTTCG CCATTACTTT CAATCACTTC TTTATACCAA TAAAATGATT TTTTCTTATA 10920

GCGATTTATA GTCAATTGAA ACAAGAGCAG GACAAAAGAG CCTCATAAAA GGTATTGCAA 10980

CTTGGTAATA CCTTTTTGAG GTGCTTTTTG ATATGAGCCC ATGTTTTCTC AATAGGATTG 11040

TACTCAGGTG AGTAGGGAGG AAGAGGTAAA AGTTTATACC CAAACTCTTC ACACAAGAGT 11100

TCTAGCTTCC CCATTCTATG GAATCTTGCA TTATCCATAA TAATAACCGA TGGTGTGGTT 11160

AATGTTGGTA AGAGAAACTT CTGAAACCAA GCTTCAAAAA AGTCGCTCGT CATCGTCTCT 11220

TCGTAAGTCA TTGGAGCGAT TAACTCACCA TTTGTTAGAC CTGCAACCAA AGAAATCCTC 11280

TGATATCTTC TTCCAGATAC TTT 11303

(2) INFORMATION FOR SEQ ID NO: 116:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3112 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

CCTTAGATTT CCACTTGCCA GAGGAATTGA TTGCCCAAAC GCCCCTTGAA AAACGTGATG 60

CCTCCAAACT CCTCATCGTC AACCGTGAGA CAGGAGAAAT GCAAGATAAA CATTTCCACT 120

CTATTATTGA TATGCTGGAA CCTGGTGATG CCCTTGTCAT GAACGACACC CGAGTTCTCC 180

CTGCCCGCCT CTATGGTCAA AAAGTGGAGA CAGGAGGTCA TGTGGAACTT CTCCTCCTTA 240

AGAACACTAG TGGAGACGAG TGGGAAGTTC TGGCTAAACC TGCCAAACGC CTCAAGGTCG 300

GTACTCGTAT CAGCTTTGGT GATGGCCGCC TCAGCGCTGT CGTTACAGAA GAATTGACCC 360

ACGGGGGACG CATTGTCCGC TTTGAATACC AAGGAATTTT CCTAGAAGTC TTGGAAAGTC 420

TGGGAGAAAT GCCTCTGCCA CCTTATATCC ACGAAAAATT AGATGACCGT GAACGTTATC 480

AAACCGTCTA CGCCAAGGAA AGTGGCTCTG CTGCAGCACC GACTGCTGGT CTTCACTTCA 540

CCAAAGAACT GCTGGCAGAA ATCCAAGCTA AGGGTGTTCA TCTAGTCTAT CTGACTCTCC 600

ATGTCGGACT CGGAACCTTT AGACCTGTTT CTGTGGATAA TCTGGACGAA CACGAAATGC 660

ACTCAGAGTT CTATCAACTT TCTGAGGAAG CTGCTGCCAC CCTTCGCTCT GTCAAAAAAA 720

ATGGTGGTCG TGTCATCGCT GTCGGAACCA CTTCTATCCG CACCTTGGAA ACTATTGGTT 780

CCAAGTTTGA TGGGCAAATC CAAGCAGATT CTGGTTGGAC CAATATCTTT ATCAAACCTG 840

GGTATGAGTG GAAGGTCGTG GATGCCTTCT CAACCAACTT CCACCTGCCA AAATCAACTC 900

TGGTCATGTT GGTTTCTGCC TTTGCAGGCC GTGAATTAGT CTTAGATGCC TACCACCATT 960

CCATCCAAGA ACACTACCGC TTCTTCAGTT TTGGTGACGC CATGTTTATT TATTGAGAAA 1020

GAATTTCTCT AAATCTTCTA ATACCAATAA ATCGCTAAGA TATTATTTCA AAGAACATCT 1080

ACAATTGAAA CTCTAGCTAG CTGTAGAAGA GGCCTAGTAC ATTGAAATTA AAATGCTTCC 1140

CCCTAGCTTC GAAAATATTG CCATAGATTG CGTTGACTCT CCAAATTGAT TCATCTATAT 1200

TTTATTTCAG CTTCCTATAC TTTCTTCGCT GTTTGTAAAT CAAAATGCAA GACACATGAG 1260

TAGCACCATA TTTGTTACTC TTATCTGTCC TCTCAAGAGA CTATTATGAG TTATTTCAGA 1320

ATCATTCACT ACTTTGACCC TGACTCTCCT TAGTCTCAAA ATCAAAGACT TATACTCTTC 1380

AAAAATCTCT TCAAACCGCG TCAACGTCAC CTTGGATTAT ATATGTGatC TGaCTTCGTC 1440

AGTTCTATCT ACAACCTCAA AGCAGTACTT TGAGCAACCT GCGACTAGTT TTCTAGTTTG 1500

CTCTTTGATT TTCATTGAGT ATTAAACAAA AAGTGAACAA ATCTGAATTC TAATGTACAG 1560

AAGACTAGGC TTGTTCACTT TTTTATAGTC GCTATAAGAT GACCTTATCT ATAGCTTTTT 1620

ATATATAATT ATATATTCAG ACATACTATT ATCAATTTTG TCGCAGGGAG GAATCTGTTA 1680

ACGCACCCAT TCACCATTAT CATTGACTCT ATAGCCATCT ATACTTGTAT TGACCGCTAA 1740

CTCACCCGAT GTATTTACAT AATACCATTT ACCACCAACT TGGAACCATT GATTGACTTT 1800

CATAGAACCG TTGCTGTTGA GGTAGTACCA TGAACTATTA ACTTGTACCC AACCTGTTGC 1860

CATGGAACCA TCAGTATTAT AAAAATACCA CATACCATTT TCTTGTTTCC AGTCTGTTGT 1920

TGGAGCAACT GCTTTAGCTG GTTCTACTGC TACATCTGTT CCTTGGTTAG ATGTAACAGA 1980

TACAGGATAC GAAGGAATAG ATGATTGCTC AGGAACAACA ACTTTTTCAG GTTCTCTCGT 2040

CCCTCTCCTT ATACGTCTTT TTACCATCTC TTTAGTAATT TGACGAGAAG TAGTTTCTTC 2100

AATTGTTCCA TCACGTTCAT CTACAGTATA GATTGTAGTA AGAGTAATTT ACCAATTTCT 2160

CCTACTTCTT CTACTTCTTG ACTTTTATCA AGAGTTGGGC CATCGAGATA TTCTGTTTCG 2220

ATTGGAATTT CTTGGACAAG AACTTGGGGC TTGGTTCTTT TTTTAACAAC TCTTGTTTGA 2280

GAGTCTTTTT TTTGACTTAA AGTACTCTCA GTTACTTGTC CACTCTTTCC ATCTACATTA 2340

TAAGTTATCG TTGTAACTGT TTTCCCATTC TTTCCTAGAG TAATCTCTTG CTCCTGTCCT 2400

GCAGAAAGGT CATTGTCTGC TTCATATTTA GTAGCAAATG GAACAAGAAC TTCTTCAACC 2460

TTGCTTTTAG CTGGAACTTT GATAACTGTA TCCGTGGCTT CTTTTCTATC AACAGTAACC 2520

TGTTCGGTAA CATAACCAGT CTCTGGATTA ACATCGTAGG TCCTTGTCGT AGTTACATAG 2580

CCATCCTCTC CATCAATTGT AACAGGATTT TCACTACGGT CTTTTGTTTC ATCTTTTTCA 2640

TAACGAATTC GCGTACTTGA AATTTTCTTG GTTACTACCT TAGGTTTAGT CGCTACTTTT 2700

ACAATAATAT CCCCATTGTC AGCGTCATCA TACTCTATTC CCTCTTCTTT ATCTCTAGTA 2760

TCATCTCTGA CATATTGAAT CCCATCAGCA GCATGAACAA AACTTGTATT CAGATTCCTC 2820

CTAAAAATAA AGTTAGCCCG ATTACCGCAG AACCAAAAAT CTTTCCGAGT TTACGTATTG 2880

CATAGCGCTT ATTAGTATTA GATTTTGCCA TTACATCCTA CTTCTAGTAT AGCATCTTTT 2940

840

CTATCAAACG TTAAACAATA TACGTTATAT ATAAAATAGA CTTAGAATGA TATATTGATT 3000

ATTGAACTAA CACTTTAACT ATATCGTAAT CAATCTCATA TATAAAGGAT TGCAGACATC 3060

TTATCTAAAT ACATGCGAAT ATATTTAGAT ACAAACATTC CAACTTGATA AT 3112

(2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4327 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

CCCAAAAAATC TCTTCAAACC ACGTCAGCTT CGCCTTGCCG TAGTATGGTT ACTGACTTCG 60

TCAGTTCTAT CCACAACCTC AAAACAGTGT TTTGAGCATC ATGCgGCTAG CTTCTTAGTT 120

TGCTCTTTGA TTTTCATTGA GTATAAAAAC AGATGAGTTT CTGTTTTCTT TTTATGGACT 180

ATAAATGTTC AGCTGAAACT ACTTTCAAGG ACATTATTAT ATAAAAGAAT TTTTTGAAAC 240

TAAAATCTAC TATATTACAC TATATTGAAA GCGTTTTAAA AATGAGGTAT AATAAATTTA 300

CTAACGCTTA TAAAAAGTGA TAGAATCTAT TTTTATGTAT ATTTAAAGAT AGATTGCTGT 360

AAAAATAGTA GTAGCTATGC GAAATAACAG ATAGAGAGAA GGGATTGAAG CTTAGAAAAG 420

GGGAATAATA TGATATTTAA GGCATTCAAG ACAAAAAAGC AGAGAAAAAG ACAAGTTGAA 480

CTACTTTTGA CAGTTTTTTT CGACAGTTTT CTGATTGATT TATTTCTTCA CTTATTTGGG 540

ATTGTCCCCT TTAAGCTGGA TAAGATTCTG ATTGTGAGCT TGATTATATT TCCCATTATT 600

TCTACAAGTA TTTATGCTTA TGAAAAGCTA TTTGAAAAAG TGTTCGATAA GGATTGAGCA 660

GGAAGTATGG TGTAAATAGC ATAGGCTGAT GTCCATCATT TGCTTATAAA GAGATATTTT 720

AGTTTAATTG CAGCGGTGTC CTGGTAGATA AACTAGATTG GCAGGAGTCT GATTGGAGAA 780

AGGAGAGGGG AAAATTGGCA CCAATTTGAG ATAGTTTGTT TAGTTCATTT TTGTCATTTA 840

AATGAACTGT AGTAAAAGAA AGTTAATAAA AGACAAACTA AGTGCATTTT CTGGAGTAAA 900

TGTCTTATTT CAGAAATCGG GATATAGATA TAGAGAGGAT CAGTATGAAT CGGAGTGTTC 960

AAGAACGTAA GTGTCGTTAT AGCATTAGGA AACTATCGGT AGGAGCGGTT TCTATGATTG 1020

TAGGAGCAGT GGTATTTGGA ACGTCTCCTG TTTTAGCTCA AGAAGGGGCA AGTGAGCAAC 1080

CTCTGGCAAA TGAAACTCAA CTTTCGGGGG AGAGCTCAAC CCTAACTGAT ACAGAAAAGA 1140

GCCAGCCTTC TTCAGAGACT GAACTTTCTG GCAATAAGCA AGAACAAGAA AGGAAAGATA 1200

AGCAAGAAGA AAAAATTCCA AGAGATTACT ATGCACGAGA TTTGGAAAAT GTCGAAACAG 1260

TGATAGAAAA AGAAGATGTT GAAACCAATG CTTCAAATGG TCAGAGAGTT GATTTATCAA 1320

GTGAACTAGA TAAACTAAAG AAACTTGAAA ACGCAACAGT TCACATGGAG TTTAAGCCAG 1380

ATGCCAAGGC CCCAGCATTC TATAATCTCT TTTCTGTGTC AAGTGCTACT AAAAAAGATG 1440

AGTACTTCAC TATGGCAGTT TACAATAATA CTGCTACTCT AGAGGGGCGT GGTTCGGATG 1500

GGAAACAGTT TTACAATAAT TACAACGATG CACCCTTAAA AGTTAAACCA GGTCAGTGGA 1560

ATTCTGTGAC TTTCACAGTT GAAAAACCGA CAGCAGAACT ACCTAAAGGC CGAGTGCGCC 1620

TCTACGTAAA CGGGGTATTA TCTCGAACAA GTCTGAGATC TGGCAATTTC ATTAAAGATA 1680

TGCCAGATGT AACGCATGTG CAAATCGGAG CAACCAAGCG TGCCAACAAT ACGGTTTGGG 1740

GGTCAAATCT ACAGATTCGG AATCTCACTG TGTATAATCG TGCTTTAACA CCAGAAGAGG 1800

TACAAAAACG TAGTCAACTT TTTAAACGCT CAGATTTAGA AAAAAAACTA CCTGAAGGAG 1860

CGGCTTTAAC AGAGAAAACG GACATATTCG AAAGCGGGCG TAACGGTAAC CCAAATAAAG 1920

ATGGAATCAA GAGTTATCGT ATTCCAGCAC TTCTCAAGAC AGATAAAGGA ACTTTGATCG 1980

CAGGTGCAGA TGAACGCCGT CTCCATTCGA GTGACTGGGG TGATATCGGT ATGGTCATCA 2040

GACGTAGTGA AGATAATGGT AAAACTTGGG GTGACCGAGT AACCATTACC AACTTACGTG 2100

ACAATCCAAA AGCTTCTGAC CCATCGATCG GTTCACCAGT GAATATCGAT ATGGTGTTGG 2160

TTCAAGATCC TGAAACCAAA CGAATCTTTT CTATCTATGA CATGTTCCCA GAAGGGAAGG 2220

GAATCTTTGG AATGTCTTCA CAAAAAGAAG AAGCCTACAA AAAAATCGAT GGAAAAACCT 2280

ATCAAATCCT CTACCGTGAA GGAGAAAAGG GAGCTTATAC CATTCGAGAA AATGGTACTG 2340

TCTATACACC AGATGGTAAG GCGACAGACT ATCGCGTTGT TGTAGATCCT GTTAAACCAG 2400

CCTATAGCGA CAAGGGTGAT CTATACAAGG GTGACCAATT ACTAGGAAAT ATCTACTTCA 2460

CAACAAACAA AACTTCTCCA TTTAGAATTG CCAAGGATAG CTATCTATGG ATGTCCTACA 2520

GTGATGACGA CGGGAAGACA TGGTCAGCTC CTCAAGATAT TACTCCGATG GTCAAAGCCG 2580

ATTGGATGAA ATTCTTGGGT GTAGGTCCTG GAACAGGAAT TGTACTTCGG AATGGGCCTC 2640

ACAAGGGACG GATTTTGATA CCGGTTTATA CGACTAATAA TGTATCTCAC TTAGATGGCT 2700

CGCAATCTTC TCGTGTCATC TATTCAGATG ATCATGGAAA AACTTGGCAT GCTGGAGAAG 2760

CGGTCAACGA TAACCGTCAG GTAGACGGTC AAAAGATCCA CTCTTCTACG ATGAACAATA 2820

GACGTGCGCA AAATACAGAA TCAACGGTGG TACAACTAAA CAATGGAGAT GTTAAACTCT 2880

TTATGCGTGG TTTGACTGGA GATCTTCAGG TTGCTACAAG TAAAGACGGA GGAGTGACTT 2940

GGGAGAAGGA TATCAAACGT TATCCACAGG TTAAAGATGT CTATGTTCAA ATGTCTGCTA 3000

TCCATACGAT GCACGAAGGA AAAGAATACA TCATCCTCAG TAATGCAGGT GGACCGAAAC 3060

GTGAAAATGG GATGGTCCAC TTGGCACGTG TCGAAGAAAA TGGTGAGTTG ACTTGGCTCA 3120

AACACAATCC AATTCAAAAA GGAGAGTTTG CCTATAATTC GCTCCAAGAA TTAGGAAATG 3180

GGGAGTATGG CATCTTGTAT GAACATACTG AAAAAGGACA AAATGCCTAT ACCCTATCAT 3240

TTAGAAAATT TAATTGGGAA TTTTTGAGCA AAAATCTGAT TTCTCCTACC GAAGCGAACT 3300

AGAGAGATGG GCAAAGGAGA GATGGGCAAA GGAGTTATTG GCTTGGAGTT CGACTCAGAA 3360

GTATTGGTCA ACAAGGCTCC AACCCTTCAA TTGGCAAATG GTAAAACAGC GACTTTCCTA 3420

ACCCAGTATG ATAGCAAGAC CTTGTTGTTT GCAGTAGATA AGGAAGATAT CGGACAGGAA 3480

ATTATTGGTA TAGCTAAAGG AAGCATCGAA AGTATGCATA ATCTTCCTGT AAATCTAGCA 3540

GGTGCCAGAG TTCCTGGCGG AGTAAATGGT AGCAAAGCAG CGGTGCATGA AGTTCCAGAA 3600

TTTACAGGGG GAGTTAATGG TACAGAGCCA GCTGTTCATG AAATCGCAGA GTATAAGGGA 3660

TCTGATTCGC TTGTAACTCT TACTACAAAA AAAGATTATA CTTACAAAGC TCCTCTTGCT 3720

CAGCAGGCAC TTCCTGAAAC AGGAAACAAG GAGAGTGACC TCCTAGCTTC ACTAGGACTA 3780

ACAGCTTTCT TCCTTGGTCT GTTTACGCTA GGGAAAAAGA GAGAACAATA AGAGAAGAAT 3840

TCTAAACATT TGATTTTGTA AAAATGGCTC TTTGTCAACT GTAGTGGGTT GAAGTCAGCT 3900

AAGCTCGAGA AAGGACAAAT TTTGTCCTTT CTTTTTTGAT ATTCAGAGCG ATAAAAATCC 3960

GTTTTTTGAA GTTTTCAAAG TTCCGAAAAC CAAAGGCATT GCGCTTGATA AGTTTGATGA 4020

GATTATTGGT CGCTTCCAAT TTGGCGTTAG AATAGTGTAG TTGAAGGGCG TTGACGATTT 4080

TCTCTTTGTC CTTTAGAAAG GTTTTAAAGA CAGTCTGAAA AAGAGGATGA ACCTGCTTTA 4140

GATTGTCCTC AATGAGTCCG AAAAATTTCT CCGGTTCCTT ATTCTGAAAG TGAAACAGCA 4200

AGAGTTGATA GAGCTGATAG TGATGTTTCA AGTCTTGTGA ATAGCTCAAA AGCTTGTTTA 4260

AAATCTCTTT ATTGGTTAAA TGCATACGAA AAGTAGGGCG ATAAAAATGT TTATCGCTGA 4320

GTTTACG 4327

(2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3521 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

CTCTGGCCCT GCCACTCCAA CGTTTTGTCA GGGTGCTTTT TTCATAAAGG AGTTCTTATG 60

TTAGATATCA AACGTATTCG TACAGATTTT GAAGCTGTCG CAGAAAAATT AGCTACACGT 120

GGTGTAGATG CTGCTGTCTT GAATGAAATG AAAGAAATCG ATGCTAAACG TCGTAACATC 180

TTGGTCAAGG TTGAAACTCT CAAAGCAGAA CGTAACACAG TTTCTGCTGA GATTGCCCAA 240

GCTAAGCGCA ACAAGGAAAA TACAGATGAC AAGATTGCTG CCATGCAAAA TCTATCTGCT 300

GAGGTTAAAG CCTTGGATGC TGAATTGGCA GAAATCGATG CTAAATTGAC AGAATTTACA 360

ACGACTCTTC CAAATATCCC AGCTGACAGC GTTCCTGTTG GGGCTGACGA AGACGACAAT 420

GTGGAAGTTC GCCGTTGGGG TACTCCACGC GAGTTTGACT TCGAACCTAA AGCTCACTGG 480

GATCTCGGTG AAGACCTTGG TATCCTTGAC TGGGAACGCG GTGGTAAGGT AACAGGCGCT 540

CGCTTCCTCT TCTATAAAGG CCTCGGTGCT CGTTTGGAAC GTGCTATCTA CAACTTTATG 600

TTGGATGAAC ATGGAAAAGA AGGCTATACT GAAGTCATCA CACCTTACAT AGTCAACCAT 660

GATTCTATGT TTGGTACTGG TCAGTATCCA AAATTTAAGG AAGATACTTT TGAACTCAGC 720

GATACCAACT TTGTCTTGAT TCCAACTGCT GAAGTTCCTC TGACAAACTA CTACCGTGAT 780

GAAATCTTAG ACGGCAAAGA TCTTCCAATC TACTTCACTG CCATGAGTCC GTCATTCCGT 840

TCTGAGGCTG GTTCTGCCGG TCGTGATACG CGTGGCTTGA TCCGTTTGCA CCAATTCCAC 900

AAGGTTGAAA TGGTCAAATT TGCCAAACCA GAAGAATCTT ACGAAGAATT GGAAAAAATG 960

ACAGCCAACG CTGAAAACAT TCTTCAAAAA CTCAACCTTC CATACCGTGT CGTTGCTCTC 1020

TCTACTGGAG ATATGGGCTT CTCAGCTGCG AAGACTTACG ACTTGGAAGT GTGGATTCCA 1080

GCACAAAACA ATTACCGTGA AATCTCAAGC TGTTCAAACA CAGAAGATTT CCAAGCCCGT 1140

CGTGCCCAAA TCCGTTACCG TGATGAAGCA GATGGCAAGG TGAAACTCCT TCATACCTTG 1200

AACGGTTCTG GACTTGCAGT TGGACGTACA GTGGCTGCAA TTCTTGAAAA TTACCAAAAT 1260

GAAGATGGTT CTGTGACCAT CCCAGAAGCA CTTCGTCCAT ACATGGGTGG AGCTGAAGTC 1320

ATCAAACCAT AAAAAATAAG GTTTAGCTAT TTCTAGCTAG ACCTTTTTTC GTAACCAAAT 1380

CAGATAAGCA CCTAGTACAA AGAATAAAAT AGTTAGGCAT ATAATGGTTT CAGCCAATAC 1440

CAGGTAATCC AGAAATGGAA GTTTCAAAAT TCCCTGAGCC ATCTTGAGCG AGGTCGCTGT 1500

GATAATGGTT GGGAAGGTGA GGGCTGAGAA GGCTGGTTGA AAACCTTGTT TTAAAATGTT 1560

GGGCAGACGA GTTAAAACAA AGAAAAAGAA GGATTGAGAA GCCAAAATCA TGACAATCAA 1620

GACCCAAGTC GGCAGGCTGG TTCCTCCTAC TCGAACTAGA GAAGCCAAGA GTAGAGAGAA 1680

AGGAGCACAG TAGATTCCTT CTTGTCCAAG CAAGGCTAGT GGGAGTGGAT GTTTCTTTAA 1740

ATCGCTATAA ATAAGGGGAT AGAGATAGAA GGTCAAGAGA AAACCAAAAC TCAAGGTCGC 1800

ATAGGCAATT TCGATAATAC CTACCAGAGG ATAGGTCAAG GCAGCCACTG CTATCCCCAC 1860

ATAGAGAACC GTCCAGCTTG GAGTGGCATG AACCCTCCGC CCTGGACAAG CAAACTTGAT 1920

GGTAAAACCA GCAATCAAGG TCAAATCCAA GAGAAATGAA AACCACCAAA TCCCTTGTGC 1980

TACCAAAGGA AGATAAGAGA ATACGCGAAA GACATAGGTC GATAAAATCA TCCCAGCCAT 2040

AGGAAAGGTT GCCATTCCTG ACAAAAGAGG GGGCTTGGTC AATTCTTGCT TGGTTTCTTT 2100

CCAATTAAAG AGATGCAGAA TTAGAAAGTA AATCCATAAA ACCAAACCAA TCAGACTAAA 2160

AAGATGGGAT AGAACCGGCA ACGTATCTAA AATAAGATTT CCAGCTCCTG CCAAACCTAG 2220

CAAACAACCT GAAAATACTA AGGGGAGTTT TTTCATCCTA ACCTCCAATA ATCATGTTAG 2280

TTTCAGTATA ACATAAAAGC GCTTAAATGA GGATTTAAAA AAACGAGTCC GCTTATTTCA 2340

GACTTCATTT TACTCAGATA TGAATTAGGC ATAAGGTTGC AATTCTGGAT TAATTGGTGT 2400

ATTAGCTAAG TTGTTGGCAT AGTTACAGAG GATTGCTAGG CTGACACCAA AAACCACATC 2460

CAAGGCATTT TGTTGAGTGT AGCCAGCTTC TAAAAACTCA GACAAGGCTT CATCTCCTAC 2520

ACGACCCTTG GTATTGATAA CTGCCAAGGT AAACTTAGCT AGGGTATCCA ATTTAGGATC 2580

TGTTTCAATT GGAGTACGAT TGCGAAGAGC TTGAATCAAG TCATCATTCA TCTGGATTTG 2640

TTTGATGGAA AAGGCTGTGT GACCTGCGAC ACAGAAGGCA CAACCATTGG TCACGGCTGC 2700

CGTGATTTGC ACCACTTCAC GCTCAACGGG TGTCAGGCTG TTGCGACGGT GGATAGATGA 2760

GACAATTTGG TAGGCTTCTA AAACAGTCGG GGCATTGGCC AAGAGACCGA TTAGGTTGGG 2820

AATATAGCCA TTGTTGTCTT TTTCTACTGT TTCAAGAATT TCTTTCACTT CTGCTGGTGC 2880

TGACTCTACT GTATGGATAG TAAATGTTGT CATAAGATAC CTCTTTTCTT ATTATTGACA 2940

CTAATATTAT TGGAAAATCT TATAAAATCC TGATTCCTAA GTTTATCTAA GATAAAGCTT 3000

TATTCTCTCA TAAGATTTTC GTTGTTATAT TAGTTTATCA CACTTCCAAT CACTTGTATA 3060

ATATATATTA TATATCAGGC TGATAAAAAT TATTTATAGG CAAAAAAATC ACACGAGCTG 3120

TGTGATTCCA TTATTTGTCA AAATACTTTT TAGTTTCAGC AATAACGACT GGCGACAAGA 3180

CCAAGAGGGC AATCAAGTTT GGCAGAGCCA TCAAGGCGTT AACGATATCT GCGATAATCC 3240

AGACCATATC CAACTCGATA AATCCTCCTA ACAAGACCAT GAGCACAAAA ACCACACGGT 3300

AGAGCCAGAT AAAGCGAACC CCAAAGAGGA ACTCAAAACA GCGTTCTCCG TAATAGTTCC 3360

AACCTAGAAT CGTTGTAAAG GCAAAAAGTA CAAGGAAGAT GGTCAAGAGA GCAGGCCCAA 3420

AGTGTGAAAA GTTTGTTGAG AAAGCTGACT GAGTCAAGGC AACCCCATTC AAGTCACCGC 3480

TCCAAACTCC AGTTACCAAG ATGGTCAAAC CAGTTAGAGT A 3521

(2) INFORMATION FOR SEQ ID NO: 119:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1968 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

AACCTGGGCA AGCAAGCTAA AAGCAATGGG ACCTGGAATC CTAATGGCAA CTGCCGCTGT 60

TGGAGGTTCC CACATTGTAT CCTCAACTCA AGCTGGCGGT TCTTACGGTT GGTCTCTACT 120

TCTCTTGGTC ATCTTAGCCA ATGTCTTTAA ATATCCATTT TTCCGTTTTG GTGCTGAATA 180

CACAGCTGAT ACTGGAAAGA CTTTGGTTGA AGGTTATGCC GAAAAAGGAA AACTCTATCT 240

CTGGATTTTC TTTATCCTCA ATGTCTTTTC GGCTATGGTC AACACGGCTG GTGTTGCCAT 300

TCTGTGCTCA GCTATCATCG CCAGTGCCTT CCCAATGATT GGACTTAGCA TTACTCAGTG 360

GTCCCTCATT CTCGTTGCAA TCATTTGGGC TATGCTACTC TTTGGAGGCT ACAAACTTTT 420

AGACGGCATG GTCAAATGGA TTATGTCTGC CTTAACCATT GCGACTGTTC TTGCAGTTAT 480

CATTGCGGCG GTCAAGCATC CAGAATACAG TTCTGATTTT GTCGAGAAGA CACCTTGGCA 540

AATGGCAGCT CTGCCCTTCA TCGTCTCCCT CCTAGGATGG ATGCCGGCTC CTATTGAAAT 600

TTCAGCCATC AATTCACTTT GGTCAGCTGA AAAGAGAAAG ACCGTCAACT TTAACACAGA 660

AGACGCTCTG TTTGACTTTA ACACTGGTTA TATTGGAACA GCTATCCTAG CCGTCTTCTT 720

TGTGGCACTG GGAGCACTGA TTCAGTATCC TACAGGGCAG GCGGTTGAAG CTGCTTCAGC 780

CAAATACATC TCTCAATTCG TGGGCATGTA TGCCTCTGTT CTTGGCGAAT GGTCCCGTTA 840

CTTGATTACC TTTATTGCCT TCCTCTGTAT CTTTGGAACA GTTATAACTG TTATCGATGG 900

CTATTCTCGC GTTAATCAGG AATCTCTCCG ACTGCTAATC AGTCAAAAAG AGGACAATCG 960

TAAATCTTTG AACATCTGGA TGACCATCAC TGCTATCATC GGTATCGTCA TTATCAAGTT 1020

CTTCGCTGGT CAGGTTTCAA CCATGCTCCG CTTTGCCATG ATTGGCTCTT TCCTGACAAC 1080

ACCTTTCTTT GCTCTTTTGA ATTACGCCTT GGTAACGCGT GAAAACAAAA ATCTTCCTTC 1140

TTGGCTCAAA CACCTTGCCA TTGCGGGATT GATTTTCCTC TTTGCTTCGC CATCTTCTTT 1200

ATCTACGCAC TCGCAATCGG AAAAGCAGGG TAAGGGACAA GCGCGAGATG AAGATAAGGT 1260

TTCATTTCAA GAGAAAATTC AGCAAATATT TCTATGATAA AAAGCATAAG AACAAGGTTT 1320

TGAAGACCTG AACTTATGCT TTTTTACGTT CTTAAAGACT GTTTATACTC AAAAAACAGT 1380

TGAACAACTT CAACCACCTC TTATAAGAAC TTTATACTAT TCGAGAATCT CTTCAAACCA 1440

CGTCAGCTCT ATCTGCAACC TCAAAGCTGT GCTTTGAGCA ACCTGCGACT AGCTTCCTAG 1500

TTTGCTCTTT GATTTTCATT GAGTATTAAT TCTCCTTTTC CAACTCATAC AAATCTGCGA 1560

TAATAGCTGC GACATGTTTG ATATCTTCCA GCATGCCTCG CATTTCAAAG TCAGCCAATA 1620

CAGGGAAGCC AAAGCGTTGA CTGTATTGCT TGGCTGTTAG GCAGTATTGG TTATTAAAGT 1680

TACGATTTCC TGACCCAACC ACACCAAAAC ACTTACTAGC ATTGTTACCA TAGGCAATAA 1740

AATCTCCCAC CGGTGTCGTC AAAATCTCAA CATCTCCGTT ATCCACGCCA TTCCCACCTT 1800

CGAGATAGGT CGGCAAAAAA GCGACATAGG GATGGTCCAT TTCATAGAAA TTTTTGCCTT 1860

CCTTGACCAA ATCCTTGATA TGAATCTTTT GAACCTCAAT CCCTTTGTAC TGGGACAAGA 1920

GATAGTCTTT CAAGCGCGTC ACAAAACTTT CAGTGTTGCC ACTCAAGG 1968

(2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7172 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

CCGCATTTTT TATCACTAGA CTCGAGACAT CTTTTGAGTG GCTCTTGCTC TCTGGTTTAA 60

TTTTCTTCCT TGCTCAAGGA CTCCTGCTAT TTCTCTTGGT CGTCCGACTC AAACATCAAT 120

TCGCTGAGAT TTATCCTCAA ATCAATAAAA AGATTCGCTT CTACTATTTA GGGGTTCTCA 180

CCATTGATTT TCTATTTTTT GTTCTCTTAG CCTTCATTAG TTCTCAGCGT TTTTCATCTC 240

TTATGCCAAT CATCACTGCT TGCCATTCTA CTTTTTATTA TATGACAGCT GACTACCTAA 300

GAGAAAACTA TCCAGACTTT TACGACAAAC ACATCTCTTT ATGGGAGTGT CTCTAAAGAA 360

AAGGAGGTTT TAGCATGAAA AAAATCATCT TCATCAAAAC CATTCAACTC CTTGTCATTG 420

ATGGAATCAT GCTGGCATTT TTGACATTTA AAAGGGGGCT TACTTGGGAC TGGATTTTGA 480

TTTATAGCGG TTGGCTCATT TTCTTTCATC CTGTGCTATT GACCTATCTT TCAAACCAAC 540

TTTGTGACCA CTTTAGTTAA CTCTATTCCC AGATTAGACC GAGATTCTGG CGTTTTGCTT 600

TACAAATTCT CCTATGGGAT AGCCTGATGA TTCTCTCCTT GGTGTCTTTA AGTGATATTC 660

CACTTTTCCT TCAGGGAACT CTCCTCATCC TAGGACATCT CATCCCTTCC TATCGCATCT 720

GCCAAAGCCT GAAAAGAGAC TTCCCCCAAG CATATCAAGA ACCGATTTCT TTTTGGAGTA 780

TTTTATGATA GATGAGAAAG ACCAAGCCGA CTGGGCTTGG TCTTTCTTAT CTCTTTTTAG 840

TATCTAGGAT AATGGTAACA GGTCCATTAT TAACCAGCTC AACCTGCATA TCTGCTCCAA 900

AGATGCCTGT CTGAACGGGC ACTTCTTGCG CTAATTTTTG ATTGAAAGCA TCATAGAAGT 960

CTGATGCCAT ATCAGGTTTA GCTGCCCCTG TAAAGgCTGG ACGATTGCCT CTCTTAGTAT 1020

CCGCAAAGAG GGTAAACTGA GAAATAGAGA GGATTTCTCC TTCAATATCT TTGACAGACA 1080

GGTTCATCTT GCCTTCTGCG TCTGAAAAAA TCCGCATATT GACCAGTTTT CTCACAGCAT 1140

AGTCCAAATC TTCCTCTTGG TCCTCTGGTC CAACACCAAC CAGCAATAAA AGTCCCTGAT 1200

TGATTTTTCC CTGAATCTGG CCTTCTATAC TCACTTGGGC TTTTTTAACC CGTTGGATAA 1260

TGATTTTCAT AATAGCCTTT CTAGTAAGAG CTAGGACAAC TAGCCGTTGG TCCGTTTGAC 1320

AGAGTAAACT TCTGGCACAC TCTTAATTTT ATCGACAACC GTGGTCAGTG TAGAGAGGTT 1380

GGCAATACCG AAGgACACAT GGATATTAGC AAACTTCATA TCCTTGGTTG GTTGGGCATT 1440

GACCGTTGAA ATATTCTTGG TTGTATTTGA AAGAACTTGC AGTACATCGT TCAACAGTCC 1500

TGTACGGTTG AGACCGTAGA TATCGATATG GGCCATATAC TCCTTATTTG AGCTAGGGTA 1560

CTGGTCTTCC CATTCCACAT CAAGGAGACG TTGCTCGTAG TTTTCTTGGG CACGCAGGTT 1620

CATACAGTCC ACACGGTGAA TAGCCACACC ACGACCCTTG GTAATGTAGC CAACAATATC 1680

GTCACCAGGC ACGGGGTTAC AACACTTAGC AATCCGCACT AGGAGACCAG AAGCACCTTC 1740

AATAACCACT CCCCCCTCAT GCTTGACCTT GAGGGTTTCT TTATTTTCAA CCTTGACCTC 1800

GCCACCTTTG ACAAGCTCCT CTGCCTCAGC TTTGGCCTTG GCACGCTCTT CCTCACGGCG 1860

TTCCTTTTCA GTCAGACGGT TAAAGACGGT AATCGCACCG ATTTCCCCAA AACCAATGGC 1920

CGCAAAGAGG GAGTCTTCTG TCTTGTAACT GGTCTTTTGC AGAACTTGAT CCATGTGGCG 1980

CTTGTCCATA AATTTATTTG CCACATAGCC ATTTTCTTGG AACTGAGCCA TCAGCATCTC 2040

ACGACCCTTG TTGACAGACA ATTCCTTATC TTGGTTTTTA AAGAACTGGC GAATCTTATT 2100

GCGCGCCTTG CTAGTCTTGA CCATATTGAG CCAGTCACGG CTAGGTCCAA AGGAGTTCGG 2160

GTTGGCGATA ATTTCAACCT GATCCCCTGT CTTTAACTTG GTTGTCAGTG GAACCATGCG 2220

GCCATTGACC TTGGCACCAG TTGCTTTTTC ACCGACCTTG GTATGGATTT CGTAGGCAAA 2280

ATCAATCGGT CCTGAATCTT TGGGAAGGGA ACGGACAGCT CCATCTGGGG TAAAAACGTA 2340

AATCTCCTCA GCCAAATAGT TTTCCTTAAC AGAGTCCACA AATTCCTTAG CATCATCAGC 2400

CTGGTCTTGG AGCTCCATCA TCTCCTTGAT CCAGTTCATT CCAATAGCTG ATTCCTTGCT 2460

GTTAACTTGC CCCTTTATAC CTTTCTTATA AGCCCAGTGA GCCGCAACCC CGTACTCAGC 2520

CACCTCGTGC ATTTCCTTGG TTCGAATCTG GAATTCAATC GGCCCTTTTG GTCCATAAAC 2580

AGTCGTATGG ATAGACTGAT AACCATTGGC CTTGCGGTTG GCGATATAGT CTTTGAAGCG 2640

ACCTGGCATC GGTTTCCAAA ATTCATGCAC GTAACCAAGC ATGGCATAAA CATCACTTTG 2700

GGTATCTAAA ATACAACGAA TAGCAATCAG ATCATAGATT TCCTCAAACC GTTTTCTCTT 2760

GTCCTGCATT TTGCGGAAAA TTGAGTAAAT ATGCTTGGGA CGACCATAAA TCTTCCCTTT 2820

CAAGTGACGT TCTGTCGTAT ACTCCTCTAA TTTTGTGACT ACCTCATCCA CCAAGGCCTC 2880

ACGCTCCCTG CGCTTTTCCT TCATCATATG GGTAATCTTG TAAAACTCCG TTGGATTGAG 2940

ATAACGGAAA GACAAGTCTT CTAATTCCCA TTTGACACTG GAAATCCCCA AACGATGGGC 3000

AAGCGGGGCA TAGATTTCCA TGGTTTCTTT GGAAATACGC TCCTGCTTGT CTTTTCGAAG 3060

ATGTTTCAGG GTCCGCATAT TGTGCAAGCG GTCAGACAGT TTGACCAAAA TAACGCGGAT 3120

GTCCTCAGAC ATGGCCATGA GCATCTTGCG ATGATTTTCC GCTAATTGCT CCTCGATCGA 3180

TTTGTACTCG ACCTTGCCAA GCTTGGTAAC TCCGTCAACA ATCATCCGCA CATCAGGACC 3240

AAACTCTCTT TCCAAATCGT CCAAAGTCGC ATCTGTATCT TCCACCACAT CATGCAAGAA 3300

TCCACAAGCT ACTGTTACAG CATCCAGCTT TAGCTTAGCT AAAATACCTG CCACTTGGAT 3360

AGGGTGAATG ATATAAGGCT CGCCTGATTT GCGATATTGA CCACTGTGGC ATTCAACAGC 3420

ATAGACCAAG GCCTTATGGA CAAAATGAAC ATCCTCTTCC GTTAAATATT CTTTGGTTAA 3480

AGCGACAACT TCTTCGCCTG TTAAATTCAC TTCTTTCGGC ATCTCTACTC TCCAATTCTT 3540

CCTACCATTT TATCACTTTT TTAAGAATAT GAAAACTAGA TTGGAACAGA ATAAGAAAAA 3600

AATAATTCAA AATTGCTTGA TAATTCTGAA TTATTGGTCC GTAATATACT ACGAAGTTAG 3660

ATTTTAAACT TAGGTGATAG AAGGAGAGAT AGAAGAACGG AAACCATATT GTAACCCAAA 3720

GACTTTCTGA CTTCCCCAAT TCCATTGAAG ATACGAAAGA TAAACGGTGG AACTCGTATC 3780

ACATACACTG GTACCTTGAC TGGATTTTGG AATTAATACT AAATGAAAAT CAAAGAGCAA 3840

ACTAGGAAAC TAGCCGCAGG TTACTCAAAG CACCGCTTTG AGGTTGCAGA TAAAGTTGAC 3900

GCGGTTTGAA GAGATTTTTG AAGAGTATAA AAATCCTCAA GATACTTTCT TCTATCCTTT 3960

AGTTTATAAG GAGAATACCT ATGAAAAAAA CTGCTATTTC TATCTTTGCT CTCCTAATGT 4020

TAGGAGTTTG CTGCCTGTTC CTATTCAGCC AGCAAAGCTA TAAAAAACAG TCGTTCAATA 4080

CTATGCTAAC GACCAGAACC TGCCCAGTAG GATAACTTAT AGTGAATATA GCGACAAATG 4140

AGAAGCCAAC TACGGTAGCA CTCTAAACAT CACGTCTATC AAACAAGCTA ATGACGGAGT 4200

TTATGCAACC TATGAAGGGC AATTGACACC TTTCCAATAT TGATAAATTG ATAACCAGCC 4260

TGTCTTCATC TAGTCATGCT GGTTTTTAAG TTCATTTTAA ATCCTTACCT ATTCTCCCTA 4320

ACTGTGCTAT ACTTAATTTA TACTCAATGA AAATCAAAGA GCAAACTAGA AAGCTAGCCG 4380

CAGGCTGTTC AAAGCACTGC TTTGAGGTTG CAGATAAAGT TGACGCGGTT TGAAGAGATT 4440

TTCGAAGAGT ATTAGTACAT TCTTTGAGAT TGGAGCTAGT ATGAAAATCC ATAAAACCGT 4500

GAATCCTGTT GCCTATGAAA ATACCTATTA TCTAGAAGGC GAAAAGCACC TCATCGTCGT 4560

CGATCCTGGT AGTCATTGGG AAGCCATTCG TCAGACAATC GAGAAGATCA ACAAACCGAT 4620

CTGTGCTATT CTCTTGACCC ACGCCCATTA TGACCATATC ATGAGTCTGG ACTTGGTTCG 4680

CGAGACGTTT GGCAATCCTC CTGTCTATAT CGCAGAGAGC GAAGCCAGCT GGCTCTACAC 4740

TCCTGTCGAT AATCTCTCCG GTCTCCCTCG CCACGATGAT ATGGCAGATG TGGTCACAAA 4800

ACCTGCAGAA CACACCTTTG TCTTTCACGA AGAATACCAA CTAGAGGAAT TTCGTTTTAA 4860

GGTTCTACCG ACCCCAGGGC ACTCTATCGG TGGTGTTTCC CTAGTCTTTC CTGATGCTCA 4920

TCTAGTCTTG ACGGGAGATG CTCTATTCCG CGAAACTATC GGACGGACCG ACCTTCCGAC 4980

TGGTAGCATG GAGCAACTCC TTCATAGTAT CCAGACCCAA CTCTTCACCC TACCAAACTA 5040

CGATGTCTAT CCAGGACATG GTCCAGCTAC TACTATCGCT CACGAAAAGG CCTTCAATCC 5100

CTTTTTCTAG CAAGATGATG ACAATCGAAA TTTAAGTAAA CTATCCAGCA AATCTTTCTA 5160

TTACAAAAGG CATCCTATCA AGGTTTTCAC ACATGATTGG ATGCCTTTTT TCTGATGACT 5220

AGATTTTTTG CATTACCAAA TAATCACGCG CTCCTCTGGT GAACGCCACA TTCCGTCTCC 5280

TTCTTTGACA TCATAGGTTG TAAAGAAATC GTCGAAGTTT GGTACTTGCA CATTGACACG 5340

GAGTTTGGCT GGTGCGTGCA CATCGACGCT AGCCAAAAGT TTCATAAATT CTGGTCGACC 5400

TTTCATGCGC CAGATGCGAC CGAAGTTGTA GAAGAACTCT TCTGCTGAGA AGTCTGCTTC 5460

TCTCTTAGCT GCTTCAAGCG CTGCTGCGAT TCCTCCCAAG TCAGCCACGT TTTCTGATAC 5520

AGTCAATTTA CCGTTAATGG TTGCTCCATA AGAATCCTGT CCATCAAATT GGTCAATGAC 5580

TTTTTGTGTT TTCTCCTTGA AGGCAGCATA GTCGCTCTCT GTCCACCAAT CCTTGAGGCT 5640

ACCATTTTCG TCAAAGGAAG CCCCGTTAGT ATCAAAGGCG TGGGAAATTT CATGGGCAAT 5700

CACTGCCCCA ATACCACCGT AGTTAGCAGA AGATGACTGA TGCAAGTCAT AGAAAGGCGC 5760

CTGTAAAATG GCCGCTGGAA AGACAATCAG GTTCTTCTGA GGATTGTAGT AGGCATTGAC 5820

CATATGAGCA GGCATGCCCC ATTCCTTATA ATCTACAGGC TGGTTCCACT TACTCCAACT 5880

GTGCTTGATT TCCACACGCG CAAAGGCTAG AGCATTCTCA AAAAGACTGG CAGTTTCATT 5940

CACTACCTTA TCCTTGTAAC GTGCAGGCAA TTCTTCTGGA TAGCCAATAT AAGGTTTGAT 6000

CACATTGAGC TTCACGATAG CCTGTTTACA GGTTTCTGGA GTGAGCCAGT CATTCTTAAG 6060

CAGACGCTCC TTATAAACAT CAATCATGGT TGCCACTTTT TTCTCCACAT CCGCCTTGGC 6120

TTCTGGAGAG AACTTCTCAC GGGCGTACCA AAGACCCAGG GCTTGCTTGA AAGGTTCTTG 6180

TGCTAGATGA TAAGCTGCTT TGACCTTATC TTTTGCCTCT GGAACTCCAG AAAGGGCACG 6240

GCTGTAGGCA CCAGACAAAA CACGGATATC CTCTGTTAAA TAGCTGGTTG AAAGATTGAC 6300

AACACTCAAA ATCAAGGTTG CTTTAAGGAG AGACCAGGCT TCCTCACTGT AGAATTGCTC 6360

TGCTGCTTGC CAGAAACGTT CCTCGTCTAC AATAACCTTG TCTGGTAATT GCCCAATAAC 6420

TGCTTTGAAG AAGTCATCCA AAGGTAGGGC AGGCGCGAAT TTCTTGAAAT CTTCGTAAGA 6480

ATATGGATGA TAGAGTTTAG CATATTCTGA ACTTTCTTCA TTAGAGAGCA CCACTGCCGC 6540

AACTCGGCGG TCCAATTCAA GTCTTTTTTC TAGCAAGTCT TCAATTTCTT CATCAGAGAA 6600

ATCATAAGCC TTGAGGAGAT TTGCGCTGCT TTCTTTCCAA AGAGTCAAGA GCTCTTCGCG 6660

CTGAGGATGT TCTTCTGCAT AGTAGGTCGT ATCTGGCAAG ATTGTGCTTG GAGCGCTAGC 6720

CCATAGAACA TTGATTCTAG CATCCATAAA GTCTGGCGAT ACACCAAAAG GAAGGAAGTT 6780

TGGTTTTCCT GCAAGCTCAA ACTCTGCTAG TTTAGCTGTA AAATCCGCAA AAGTCTCCAA 6840

TTCTTGGAAT TCTTTAAGGA GTGGTAAGAC AGGTGTGATA CCGTCAGCTT CTCTCTTGTC 6900

AAAATCACGA ACTAGGCGGT GGTATTTGAC AAAGTTTTCC AAGATAGCAT CCTCAGGCAC 6960

TTCTTCACCT GCTAACCACT TGTCTGTTGT CGCCAGCATC AGGTCTTCAA TTTCCTGGTC 7020

TAAATCAACA AAACCTCCTG TTTGAGACTT ATCTGCTGGG ATTTCAGCTG TCTGTTGCCA 7080

TTCTCCATTG ATAGCATCAT AAAAATCATC TTGATAACGT GTCATCTTGT TCTCGCTTTC 7140

ATTTGTATTT GCATTTATCT TAACAAAAAT CG 7172

(2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4518 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

CGGGAAGTTA TGCGATCTAG ACTTCGTTCC TGTACAGCTA CTTTCTCAGG TGGTCTTGTT 60

GTTTGTATGA GTTTGTTTAG AGAGGATCTT TCTATGTCTT TCTTTCTTAT TTTTGTTTTA 120

TATGCTTTTC TGATTTCTTA TCTAATTTAT GGTTATTTCA GACTAAAAAG GAAATACCGA 180

GTAGATGAAT AGCAAGGTTC TAGGTCTTCA GATTGATTTT TAGCACTCTT GATAAAAGAG 240

TGCTAATTTT TTGAGTTTTT GTCTTGACAT TCTCTTCTAA GGGTGTATAA TAGAATCATG 300

AGTTAGCACT TGGATGCATT GAGTGCTAAT TGATCAGACA GAGAGGAGTG ATGAGATGGT 360

TACAGAGCGT CAGCAGGATA TTTTAAATCT GATTATTGAC ATCTTTACCA AAACGCACGA 420

ACCTGTCGGA TCAAAAGCCT TGCAAGAGTC TATTAACTCT AGCAGTGCAA CCATTCGTAA 480

TGACATGGCG GAACTAGAAA AACAAGGGTT GCTTGAGAAG GCTCATACTT CAAGTGGTCG 540

GATGCCAAGT GTTGCTGGTT TTCAGTACTA TGTGAAACAC TCACTGGATT TTGACCGGCT 600

GGCTGAAAAT GAGGTATATG AGATTGTCAA AGCCTTTGAT CAGGAATTCT TCAAATTGGA 660

GGATATTCTG CAAGAGGCTG CTAACTTACT AACAGACCTG AGTGGCTGTA CGGTAGTGGC 720

ACTGGATGTT GAGCCGAGCA GGCAACGTTT GACAGCCTTT GATATCGTTG TTTTGGGGCA 780

ACATACAGCC TTGGCGGTAT TTACCCTAGA CGAGTCGCGA ACGGTTACTA GTCAGTTTCT 840

GATTCCAAGG AACTTCTTGC AGGAGGATTT GCTGAAACTG AAGAGCATCA TTCAGGAACG 900

TTTCCTCGGT CACACCGTTT TAGATATTCA CTACAAGATT CGGACGGAGA TTCCGCAGAT 960

TATCCAGCGT TACTTTACAA CAACGGATAA TGTCATCGAT CTCTTTGAAC ACATCTTTAA 1020

GGAAATGTTC AACGAAAACA TTGTGATGGC GGGCAAGGTC CATCTCTTGA ATTTTGCCAA 1080

TCTAGCAGCC TATCAGTTCT TTGACCAACC GCAAAAGGTG GCCTTGGAGA TTCGTGAGGG 1140

GTTGCGTGAG GATCAGATGC AAAATGTTCG TGTTGCAGAC GGTCAAGAGT CCTGTTTAGC 1200

TGACCTAGCG GTAATCAGTA GTAAGTTCCT CATTCCTTAT CGGGGAGTTG GAATTCTAGC 1260

CATTATCGGT CCAGTTAATC TGGATTACCA ACAGCTAATC AATCAAGTCA ATGTGGTCAA 1320

CCGTGTTTTG ACCATGAAGT TGACAGATTT TTACCGCTAC CTCAGCAGTA ATCATTACGA 1380

AGTACATTAA GATTGAAATC ATTAAAGGAG GCGAACATGG CCCAAGATAT AAAAAATGAA 1440

GAAGTAGAAG AAGTTCAAGA AGAGGAAGTT GTGAAAACAG CTGAAGAAAC AACTCCTGAA 1500

AAGTCTGAGT TGGACTTGGC AAATGAACGT GCAGATGAGT TCGAAAACAA ATATCTTCGC 1560

GCTCATGCAG AAATGCAAAA TATCCAACGC CGTGCCAATG AAGAACGTCA AAACTTGCAA 1620

CGTTATCGTA GCCAGGACTT GGCAAAAGCA ATCTTACCAT CTCTTGACAA CCTTGAGCGT 1680

GCACTTGCAG TTGAAGGTTT GACAGATGAT GTGAAGAAGG GCTTGGGGAT GGTGCAAGAA 1740

AGCTTGATTC ACGCTTTGAA AGAAGAAGGA ATTGAAGAAA TCGCAGCAGA TGGCGAATTT 1800

GACCATAACT ACCATATGGC CATCCAAACT CTCCCAGCAG ACGATGAACA CCCAGTAGAT 1860

ACCATCGCTC AAGTCTTTCA AAAAGGCTAC AAACTCCATG ACCGCATCCT ACGCCCAGCA 1920

ATGGTAGTGG TGTATAACTA AGATATAAAG CCCGTAAAAA GCTCGCAGTA AAAATAGGAG 1980

ATTGACGAAG TGTTCGATGA ACACAAGAAA ATCTATCTTT TTTACTCAGA GCTTAGGGCG 2040

TGTTCGATTC GGCAATTCTG ACGGTAGCTA AAGCAACTCG TCAGAAAACG GCAATCGCTA 2100

TGGCGTTTGC CTAGCTTCCT TACTAACTCG TCGTCGAAAT AAAATCGATT TCGACTCCTC 2160

GTGTCGCAAT TTACATAATA GAAAACTTGT CCGAAACGAC AATAAACTAT GAAGAAAGAT 2220

AAAATATGTT TGGCTTTGTA ATAGTGAGCG AAGCGAACCA AACACGATAC TCTTCGCCGT 2280

GGCGCTATTT GCGCAAATTT TGAGACCTTA GGCTCAAAGT TTAGTCAAAG AGATTGACGA 2340

AGTCAAGCTC TGACGGCGTC GCCACTGTCG CCACTTAAGA AGAGTATCAA AAAGAAAAAT 2400

AGAAAATTAA CTAACAAGGA GAAAAACACA TGTCTAAAAT TATCGGTATT GACTTAGGTA 2460

CAACAAACTC AGCAGTTGCA GTTCTTGAAG GAACTGAAAG CAAAATCATC GCAAACCCAG 2520

AAGGAAACCG CACAACTCCA TCTGTAGTCT CATTCAAAAA CGGAGAAATC ATCGTTGGTG 2580

ATGCTGCAAA ACGTCAAGCA GTTACAAACC CAGATACAGT TATCTCTATC AAATCTAAGA 2640

TGGGAACTTC TGAAAAAGTT TCTGCAAATG GAAAAGAATA CACTCCACAA GAAATCTCAG 2700

CTATGATCCT TCAATACTTG AAAGGCTACG CTGAAGACTA CCTTGGTGAG AAAGTAACCA 2760

AAGCTGTTAT CACAGTTCCG GCTTACTTCA ACGACGCTCA ACGTCAAGCA ACAAAAGACG 2820

CTGGTAAAAT TGCTGGTCTT GAAGTAGAAC GTATTGTTAA CGAACCAACT GCAGCAGCTC 2880

TTGCTTATGG TTTGGACAAG ACTGACAAAG AAGAAAAAAT CTTGGTATTT GACCTTGGTG 2940

GTGGTACATT CGACGTCTCT ATCCTTGAAT TGGGTGACGG TGTCTTCGAC GTATTGTCAA 3000

CTGCAGGGGA CAACAAACTT GGTGGTGACG ACTTTGACCA AAAAATCATT GACCACTTGG 3060

TAGCAGAATT CAAGAAAGAA AACGGTATCG ACTTGTCTAC TGACAAGATG GCAATGCAAC 3120

GTTTGAAAGA TGCGGCTGAA AAAGCGAAGA AAGACCTTTC TGGTGTAACT TCAACACAAA 3180

TCAGCTTGCC ATTTATCACT GCAGGTGAGG CTGGACCTCT TCACTTGGAA ATGACTTTGA 3240

CTCGTGCGAA ATTTGACGAT TTGACTCGTG ACCTTGTTGA ACGTACAAAA GTTCCAGTTC 3300

GTCAAGCCCT TTCAGATGCA GGTTTGAGCT TGTCAGAAAT CGACGAAGTT ATCCTTGTTG 3360

GTGGTTCAAC TCGTATCCCT GCCGTTGTTG AAGCTGTTAA AGCTGAAACT GGTAAAGAAC 3420

CAAACAAATC AGTAAACCCT GATGAAGTAG TTGCTATGGG TGCGGCTATC CAAGGTGGTG 3480

TGATTACTGG TGATGTCAAG GACGTTGTCC TTCTTGATGT AACGCCATTG TCACTTGGTA 3540

TCGAAACAAT GGGTGGAGTA TTTACAAAAC TTATCGATCG CAACACTACA ATCCCAACAT 3600

CTAAATCACA AGTCTTCTCA ACAGCAGCAG ACAACCAACC AGCCGTTGAT ATCCACGTTC 3660

TTCAAGGTGA ACGCCCAATG GCAGCAGATA ACAAGACTCT TGGACGCTTC CAATTGACTG 3720

ATATCCCAGC TGCACCTCGT GGAATTCCTC AAATCGAAGT AACATTTGAC ATCGACAAGA 3780

ACGGTATCGT GTCTGTTAAG GCCAAAGACC TTGGAACTCA AAAAGAACAA ACTATTGTCA 3840

TCCAATCGAA CTCAGGTTTG ACTGACGAAG AAATCGACCG CATGATGAAA GATGCAGAAG 3900

CAAACGCTGA AGCCGATAAG AAACGTAAAG AAGAAGTAGA CCTTCGTAAT GAAGTAGACC 3960

AAGCAATCTT TGCGACTGAA AAGACAATCA AGGAAACTGA AGGTAAAGGC TTCGACGCAG 4020

AACGTGACGC TGCCCAAGCT GCCCTTGATG ACCTTAAGAA AGCTCAAGAA GACAACAACT 4080

TGGACGACAT GAAAACAAAA CTTGAAGCAT TGAACGAAAA AGCTCAAGGA CTTGCTGTTA 4140

AACTCTACGA ACAAGCCGCA GCAGCGCAAC AAGCTCAAGA AGGAGCAGAA GGCGCACAAG 4200

CAACAGGGAA CGCAGGCGAT GACGTCGTAG ACGGAGAGTT TACGGAAAAG TAAGATGAGT 4260

GTATTGGATG AAGAGTATCT AAAAAATACA CGAAAAGTTT ATAATGATTT TTGTAATCAA 4320

GCTGATAACT ATAGAACATC AAAAGATTTT ATTGATAATA TTCCAATAGA ATATTTAGCT 4380

AGATATAGAG AATTATATTA GCTGAACATG ATAGTTGTAT CAAAAATGAT GAAGCGGTAA 4440

GGAATTTTGT TACCTCAGTA TTGTTGTCTG CATTTGTATC GGCGATGGTA CCGTATCTGA 4500

CGAACGTTCA GCTTATAT 4518

(2) INFORMATION FOR SEQ ID NO: 122:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8145 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

TGCTATTTTC GATTCCCTTG GGCGTTTTGA TTGCCTTTGC CTTGCAAGTC CATTGGAAGC 60

CCCTCCATTA TCTGATTAAC ATTTACATCT GGGTTATGCG AGGAACCCCC TTACTCTTGC 120

AACTGATTTT TATCTATTAT GTGCTCCCAA GTATTGGGAT TCGTTTAGAC CGCCTTCCTG 180

CAGCTATTAT TGCCTTTGTT CTCAACTATG CAGCTTACTT TGCAGAAATT TTCCGTGGGG 240

GAATTGACAC TATTCCAAGA GGACAGTATG AGGCCGCCAA GGTCTTGAAG TTTAGCCCTT 300

TTGACAGAGT GCGCTATATT ATCTTGCCCC AAGTGACCAA GATCGTTCTT CCTAGTGTCT 360

TTAATGAAGT TATGAGTTTG GTCAAGGATA CTTCTTTGGT CTATGCTCTC GGAATTTCAG 420

ACCTTATCTT GGCTAGTCGA ACAGCTGCTA ACCGCGATGC TAGTCTAGTT CCTATGTTCT 480

TGGCAGGAGC CATTTATTTG ATTTTGATTG GGATTGTGAC AATTATTTCC AAAAAAGTTG 540

AGAAGAAGTA TAGTTATTAT AGATAGGAGG CTGCCATGTT AGAATTACGA AATATCAATA 600

AAGTCTTTGG AGACAAACAA ATCCTGTCTA ATTTCAGTCT AAGTATTCCT GAAAAGCAAA 660

TCCTGGCTAT CGTTGGACCT TCTGGTGGAG GTAAGACAAC TCTTTTACGT ATGCTTGCAG 720

GTCTTGAAAC CATTGATTCA GGGCAAATCT TTTATAATGG ACAACCTTTA GAGCTGGATG 780

AATTGCAGAA GCGCAATCTA CTGGGATTTG TCTTCCAAGA TTTTCAACTA TTTCCTCATC 840

TATCAGTTCT GGAAAATTTG ACTTTATCGC CTGTGAAGAC CATGGGAATG AAGCAGGAAG 900

AGGCTGAGAA GAAGGCGAGT GGACTCTTGG AACAGTTAGG ACTAGGAGGA CACGCAGAGG 960

CCTATCCTTT CTCACTATCT GGTGGGCAAA AGCAGCGGGT GGCTTTGGCG CGTGCTATGA 1020

TGATTGACCC AGAAATCATT GGCTACGATG AACCAACTTC TGCCCTGGAT CCAGAATTAC 1080

GTTTGGAAGT GGAGAAGCTA ATCTTGCAAA ATAGGGAACT TGGGATGACC CAGATTGTGG 1140

TTACCCATGA TTTGCAGTTT GCTGAAAATA TCGCAGATGT ATTATTGAAA GTAGAACCTA 1200

AATAGGAGGA AAAATGGATG AAAAAATGGA TGCTTGTATT AGTCAGTCTG ATGACTGCTT 1260

TGTTCTTAGT AGCTTGTGGG AAAAATTCTA GCGAAACTAG TGGAGATAAT TGGTCAAAGT 1320

ACCAGTCTAA CAAGTCTATT ACTATTGGAT TTGATAGTAC TTTTGTTCCA ATGGGATTTG 1380

CTCAGAAAGA TGGTTCTTAT GCAGGATTTG ATATTGATTT AGCTACAGCT GTTTTTGAAA 1440

AATACGGAAT CACGGTAAAT TGGCAACCGA TTGATTGGGA TTTGAAAGAA GCTGAATTGA 1500

CAAAAGGAAC GATTGATCTG ATTTGGAATG GCTATTCCGC TACAGACGAA CGCCGTGAAA 1560

AGGTGGCTTT CAGTAACTCA TATATGAAGA ATGAGCAGGT ATTGGTTACG AAGAAATCAT 1620

CTGGTATCAC GACTGCAAAG GATATGACTG GAAAGACATT AGGAGCTCAA GCTGGTTCAT 1680

CTGGTTATGC GGACTTTGAA GCAAATCCAG AAATTTTGAA GAATATTGTC GCTAATAAGG 1740

AAGCGAATCA ATACCAAACC TTTAATGAAG CCTTGATTGA TTTGAAAAAC GATCGAATTG 1800

ATGGTCTATT GATTGACCGT GTCTATGCAA ACTATTATTT AGAAGCAGAA GGTGTTTTAA 1860

ACGATTATAA TGTCTTTACA GTTGGACTAG AAACAGAAGC TTTTGCGGTT GGAGCCCGTA 1920

AGGAAGATAC AAACTTGGTT AAGAAGATAA ATGAAGCTTT TTCTAGTCTT TACAAGGACG 1980

GCAAGTTCCA AGAAATCAGC CAAAAATGGT TTGGAGAAGA TGTAGCAACC AAAGAAGTAA 2040

AAGAAGGACA GTAAGATAAA ATAGTGGCTG AAACTGCGTT TTGATTAGCA AAACGTAGTT 2100

TTTTTTGTAA TCTAGGAAAA CGATAATAGC GATTGAATAT GGATAATTGA ATATGGAATA 2160

GCCCACTGTG ATTTCTAAAA CATTGTTAAA AATTGATTTG ACTTCCAAAA TTAAAATGTT 2220

CTGTAATGAA ATACTGATGT AACTGTTTTA GGAACAATAA AACGCATAAT ATCAAGGTTT 2280

TTGCACCTTA CATTATGCGT TTTTGTGATT TTAAGACTTG TTAGCTGATT TTTTACAATC 2340

CTGCGAAATC TTTGATTTCT TGTGCTGACA TTGAAGAGTC GCAACGGACG TTGATTTGTC 2400

CATCTGTAAT ATGAACAAAA CCTGGTACAG TTGGGATTCC ATAGCGTGAG CGGAATGCTT 2460

GCAAATCATT GAGTTGGCTT GGTTCTTCAC TATTGATGAA GTAAATGTGA GCTTTGGTTT 2520

CAGCTACGAC ACCTGACAAT GTACCTGCAA ATTTACGGCA GTAAGGGCAA GTTTTGCGAC 2580

CGATAAAGAA GGTTGCAGTT TCTTTTTTAT CAAGAGCTTC TTGCGCACGC ACAACTGTAG 2640

TGACTTCAAG GTCTTTGATG TTATCTAAAA ATTGTTCCAT GAGATTACCT CGCTTTCATT 2700

GATAAGTCTA GTATGCCATA AAGTTTCTAA AATTGCTTAG ATTTGATACG AAAAAAGATG 2760

AGGTTGGTTG GTCTCATCTT TTATAGGTCT TTATTTTACA AATGCATTGA TTTCTGCTTC 2820

GATGTTAGCA ATCTTAGCTT GTGATTCTTC GTTGGTTTCC CCTACAACTG CAATGTAGAA 2880

CTTGATTTTT GGTTCTGTAC CTGAAGGGCG AACGGCAATC CATGAACCGT CAGCAAGTGT 2940

GTATTTCAAC ACATCACTTG GAGGAGTTGT CAAGTTTGTA ACAGTACCGT CAGCAACAGT 3000

AGCAGTTTGT GCCTTGAAGT CTTCTACGAC AGTGATAGCT GTTGCGTTCC ATTCTGTTGG 3060

AGCATTGTTG CGGAATTTAG CCATAATCGC TTTGATTTGT TCAGCACCAT CGACACCTGA 3120

AAGAGTAACA GAGATTGTTT TTTCTGCGTA GTAGCCATAT TCTTTATAGA TTTCTTCGAT 3180

ACCGTCAGCA AGTGTCAAAC CACGAGAACG GTAGTAGGCA GCAAGTTCAG CAACTACAAG 3240

AACGGCTTGG ATGGCATCTT TATCACGTAC AAATGGTTTA ATCAAGTAAC CGAAGCTTTC 3300

TTCAAATCCC ATCATGTAAG TGTGGTTGTG TTTTTCTTCG AATTCTTGGA TTTTTTCAGC 3360

GATAAATTTG AAACCTGTCA AGACGTTGAA CATAGTTGCG CCGTAGCTTT CAGCAATCTT 3420

CGTTACCAAG TCAGTTGAAA CGATAGATTT GCAGAGAGCG GCATTTTCAG GAAGAGTTCC 3480

AGCGTTTTTG TGAGCTTCCA AGATGTATTT AGCCATGATA GCACCGATTT GGTTACCTGA 3540

AAGGTTGAGG TAGCTACCAT CTTTTTGAAG AACTTCAACA CCAACACGGT CAGCGTCTGG 3600

GTCAGTTGCG ACAAGAACAT CTGCACCAAC TTGACGACCA AGTTCTTCAG CAAGGGCAAA 3660

GGCTGCTTGG CTTTCTGGGT TTGGAGATGT TACAGTTGAA AAGTCTGGGT CAGCAGTTGC 3720

TTGCGCTTCA ACAACTTGAA CAGAGTCAAA TCCTGCTTGG GCAAGAGCAC GACGAGCCAA 3780

CATTTCACCA GTACCATGAA GTGGTGTGTA GACAATCTTC ATGTCTTTAC CAAATTCTTC 3840

AATCAAGGCT GGGTTGATGT TTATGTCCTT AACCTCTTTA AGGTATTCTA TGTCAACAGC 3900

TTCGCCGATA ACTTCAATCA AGCCAGAAGC TTTTTCAGTT TCCACATCAG CAACTTCAAC 3960

TGCAAATGGG TTTTCGATTG CACGGATATA AGTAGTCAAA GCGTCCGCAT CGTGTGGAGG 4020

CATTTGTCCA CCGTCTTCAC CGTAAACCTT GTAACCGTTA AATGGAGCAG GGTTGTGGCT 4080

GGCTGTGACC ATGATACCTG CGAAACAGTT GAGATGACGA ACTGCAAATG ATAGTTCTGG 4140

AGTCGGACGA AGGCTTTCAA ATACGTAAGA TTTGATGCCG TGTTTAGCAA GAACTGCCGC 4200

AGATTCAAAG GCAAACTCAG GTGAGAAGTG ACGGCTATCG TAGGCAATTG CTACACCGCG 4260

TTCTTTCTCG TTTCCACCTT TTGACTCAAT CAAACGAGCC AATCCTTCAG TAGCTTGGCG 4320

AACAACGTAG ATGTTGATAC GGTTTGTACC AGCACCAACC AAGCCACGCA TACCTGCAGT 4380

ACCAAATTCA AGATTTGTAT AGAAGGCATC TTCCTTAGTT TTTTCGTCCA TATTTTCCAA 4440

ATCTTGACGA AGGTAGTCAC GAAGCTCCAC AAAATCAACC CATTTCTGGT AATTTTCTTG 4500

GTAAGACATT CAAATTCTCC TTTATTTTTA AAACATTTAA TCAGTTTAAT TATATCATTT 4560

TTTTTAGTTT TAGTAAAACC TTATCTGCTT CGAACATCTC TTCAAACCAG GTCAGATTGA 4620

ATTTTGGGGT TATATGATGT TGAGGCTAGG AAAAATTCAA TTTCAGTAAA AAAAGTAAGT 4680

CTTCTCATAA CAAAACATTG ATATAGTTAC TTAGTTTTAA ACAAGCATAT TATAATAAAG 4740

CTATGGCATA TAGTACTGAT TTTAAACAGC GAGCATTAGA TTACATCAAA GAGGGGCACA 4800

GCCATGTCGA GGCAGCCAAG TTTTTTGGTG TTGGCGTCAG AACTCTCTTC ACGTGGGAAA 4860

AGAAAGACGT GAACAAGAAC ACATAGAGAG GAAAAAGCGA GTCGTCAAAA ACCGAAAGAT 4920

TCCTTTAGAG GAATTGAAAG CCTTTGTAGA GGCTCATCCA GATGCTTTTT TACGGGAAAT 4980

TGCGGCACAT TTTGATTGTG CTGTTCCTTC AGTATGGGCA GCTTTAAAGC AGATTAAGGT 5040

CACTTTAAAA AAAGATGACG AGCTTTAAGG AACAAGACCC AGAAAAGTAG CCTTATTTCT 5100

TAAGAATTTT AATAGTTTAA AGCACCTAGC ACCTGTTTAT ATTGATGAAA CAGGAATCGA 5160

CCGCTATCTC TATCGTCCTT ATGCAGGGGC TCCTAGAGGG GAGAAAGTCT ATGAAAAGAT 5220

TAGCGGACGT CGTTTTGAGC GAACTTCAAT TGTTGCAGGA CAAGTAGACG GAGAGTTTAT 5280

AGCTCCCATG ATTTACAAGA AAAGCATGAC AAGCGATTTC TTTGTGGAGT GGTTCAAAAC 5340

GCAACTCCTA CCTGCTTTGA AGACACCTCA TGTTATTGTC ATGGGCAATG CTGGTTTTCA 5400

TCCCAAGAAC ATTTTGGATG AACTCTGCAT CCAAGATAAA CACTTTTTCT TACCTCTACC 5460

ACCTTATTCA CCGGATTTGA ATCCTATTGA GCAAGCTTGG GCTATCTTGA AAAAGAAAGT 5520

GACGGATGTA TTAAGGGAAG TTCCAACTAT TTTTGAATGT TTGGAATGCT TTTTTAAAAC 5580

TAGATGACTA TAACGGTTCT AAAGGAACCT ATCGAGTAGT CATTAAAACT AAGGATACTG 5640

CTGGTTAAGA GAAGACGGTA TACAATCAAA CCATTCACCG TGTAGCCGAA ATCGTTCAGA 5700

ATGAAGACTT GTATCAGAAT GAAGACTTGT ATAAGAAAGG TTTGAATGTT GAACTTGCGC 5760

ACCAACAAAT TAAGGGATTT TTTGAAGCAG AGTTTAAAAA TCGTATTAAT GGAGTTCTTA 5820

ATACTAAAAT AAAAAATAGT ACATTAAATC GTGTAAATAA AAAAACTATA CACCAGAGCA 5880

ACAAAAACTC CATGATCAAT TTGAAGCAGA AGCAACGGAA GATGCTAAAA AACAAGGCGA 5940

TATTGTGTTG AATGTTGACC AGGATTTCAT GAGCATATCT AAGTCTAATA AAAGTGGTTC 6000

AGACTGGAAG AAAACTTTCA CAGTGAGGAT AACCAATAGG CTAGCAAATG ACTTGAATAA 6060

TGTCTTGAAA CAGGTTGATA AAGATACTCC TAATACCCCA ACTTGGCTAA ACTCAGCTGC 6120

TTCTAAAGCT AAAGATGATG ACAGAGTATA TAAACTACTG AAGACTCTTA TACCAGGAGA 6180

AAATTACCTA TCATGTTAAG GATAATCAGC TAGAAGTAGA AACAGATAAA TACACATATA 6240

CTGCCGCTAG AAATGGTAGT AAGGAAGTTG GTATTCAAGA GTCAGATATA GCAGCAACTC 6300

TAAGTGCCGA TGAATATAAT TCTAATCGCC AAACTTTTGA GAGAGAATAC AAATACAAAA 6360

GCAAATGCCC TTAATAATGG TTGGGCTAGA TCTGGTTCTG AAGAGTTCAA AAAGTTCTCC 6420

CACTTTGTAG GGGTAGACAA AGGGATTGTG CGAACGAATG TACTGACTGG TAAAAAACTA 6480

TCTGATAAGA TTAGGAAAGA AGTGGGCTCT GGAGATAGCA AACTAGGAAA AGGCGGCTAT 6540

TTCTCTACTG GGGATGTTCT ATTAGGAAAA GATGTTGTTT CTTATACCGT ACAAGTATTT 6600

TCAGAGAATA ATGAAAGAGT AGGAGTAAAC ACTCAAAGTC ACCGTGTTCA GTATAATCTC 6660

CCAATTCTAG CTGACTTTTC AGTCATCCAA GATACTGTGG AACCATCACG AACCGTTGTT 6720

GAAAAAATCA TTCCAAAACT AAATATTCCC GAAGAAGAGA AAGGGAAAAT AACCGAAGAA 6780

ATCAAGAAAA AGAAAAAAAC CTCAGAATTG GCAGAACTAA TCTCAGAAAA TGTGAAAGTT 6840

CGCTATGTTG ATGAACAAGG GCGTTTGCTA TCATTGAAAA ATGATACTGG AATTGGAGAA 6900

AAAGAAAGTG ACGGAACCTA CATTACCAAT AAAAAACAAC TGATTGGTAC CAGCTATAAT 6960

GTCACAGATA AAAAACTCAG TAGCATGACT ACTACTGACG GAAAATATTA TACTTTTAAA 7020

GAAGCAGATA CAAATTCTGC AAGTTTAACT GGGAATATTG TAAGCGAAGG TAGAACAGTG 7080

ACCTTAGTTT ATAGAGAAAG CGAAGCGCCA ACCACTGCTA CAGTAACAGC CAATTACTAT 7140

AAAGAAGGTA GGCAAGAGAA GTTGGTAGAG TCTGTTATAA AAGCTGATTT AGCGATAGGT 7200

TCTGAGTATA CCACAGAATC AAAAACTATT GAAGGGAAAA CAACAACTGA GGACAAAGAA 7260

GACCGAGTTA TCACAAGGAA AACAACATAC ACCTTGGTAG CAACTCCTGA AAATGCGTAC 7320

CAGAAGACGG TGCAACAGTT GACTATTACT ACCGTGAGAA TGTTGAGGAA ACAGTGGTTC 7380

CCAAAACAGC AACCTCTACT GAGACGAAGA CTATAACGCG TATCATTCAT TACGTTGATA 7440

AAGTTACGAA CCAAAATGTA AAAGAAGATG TTGTTCAACC TGTAACCTTA AGCCGTACAA 7500

AAACTGAGAA CAAGGTCACG GGAGTTGTAA CCTACGGTGA ATGGACAACA GGAAACTGGG 7560

ACGAGGTTAT ATCTGGTAAG ATTGACAAGT ACAAAGATCC AGATATTCCA ACAGTTGAAT 7620

CACAAGAAGT TACGTCAGAC TCTAGTGATA AAGAAATAAC GGTAAGGTAT GACCGTTTAT 7680

CAACACCAGA AAAACCAATC CCACAACCAA ATCCAGAGCA TCCAAGTGTT CCGACACCAA 7740

ACCCAGAACT ACCAAATCAA GAGACTCCAA CACCAGATAA ACCAACTCCA GAACCAGGTA 7800

CTCCAAAAAC TGAAACTCCA GTGAATCCAG ACCCAGAAGT TCCGACTTAT GAGACAGGTA 7860

AGAGAGAGGA ATTGCCAAAC ACAGGTACAG AAGCTAATGC TACCTTGGCT AGTGCTGGTA 7920

TCATGACCTT GTTAGCTGGT CTAGGATTAG GATTTTTCAA GAAAAAAGAA GATGAAAAAT 7980

AATAGATTTT AGAATCTAGG AACCAGGAAA AGCTCACAGA TGTGGGCTTT TTTCCTGGTT 8040

TTGAGAACGA GGTCTTTCGT AAAGAATAAA AACGCTTACA AGTCTGTTGA ACTGGGAAAC 8100

TATGAATCCT ATTTTTTTAA AAATATTTCC AGAAATCAGT TGCGG 8145

(2) INFORMATION FOR SEQ ID NO: 123:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8697 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

CGGTACCGGG AACGATACTT AGTCTAATTT TGCACCTTTT CCATGTATGG TAAAGGTTTT 60

TCTTTTTTTA AAAAGGAAAA CGAGAAGAGG AGGTTCTTAT GAAAGCAAGC ATTGCCTTGC 120

AAGTTTTACC CCTAGTACAG GGGATTGATC GGATAGCTGT TATTGATCAG GTCATTGCTT 180

ATCTGCAwAC TCAAGAAGTG ACGATGGTAG TGACACCATT TGAAACGGTC TTGGAAGGGG 240

AGTTTGATGA GCTTATGCGC ATTCTAAAAG AAGCGCTGGA AGTGGCAGGG CAGGAGGCAG 300

ACAATGTCTT TGCCAATGTC AAAATAAATG TAGGAGAGAT TTTAAGTATT GATGAGAAAC 360

TTGAGAAGTA TACTGAGACG ACACATTAGT CTATTGGGCT TTCTCGGAGT ATTGTCAATC 420

TGGCAGTTAG CAGGTTTTCT TAAACTTCTC CCCAAGTTTA TCCTGCCGAC ACCTCTTGAA 480

ATTCTCCAGC CCTTTGTTCG TGACAGAGAA TTTCTCTGGC ACCATAGCTG GGCGACCTTG 540

AGAGTGGCTT TACTGGGGCT GATTTTGGGA GTTTTGATTG CCTGTCTTAT GGCTGTGCTC 600

ATGGATAGTT TGACTTGGCT CAATGACCTG ATTTACCCTA TGATGGTGGT CATTCAGACC 660

ATTCCGACCA TTGCCATAGC TCCTATCCTG GTCTTGTGGC TAGGTTATGG GATTTTGCCC 720

AAGATTGTCT TGATTATCTT AACGACAACC TTTCCCATCA TCGTTAGTAT TTTGGACGGT 780

TTTAGGCATT GCGACAAGGA TATGCTGACC TTGTTTAGTC TGATGCGGGC CAAGCCTTGG 840

CAAATCCTGT GGCATTTTAA AATCCCAGTT AGCCTGCCTT ACTTTTATGC AGGTCTGAGG 900

GTCAGTGTCT CCTACGCCTT TATCACAACT GTGGTATCTG AGTGGTTGGG AGGTTTTGAA 960

GGTCTTGGTG TTTATATGAT TCAGTCTAAA AAACTGTTTC AGTATGATAC CATGTTTGCC 1020

ATTATTATTC TGGTGTCGAT TATCAGTCTT TTGGGTATGA AGCTGGTCGA TATCAGTGAA 1080

AAATATGTGA TTAAATGGAA ACGTTCGTAG AATTAGAATG TTTCTGAAAA AGAAAAGAGG 1140

AAATCAAAAT GAAGAAAACA TGGAAAGTGT TTTTAACGCT TGTAACAGCT CTTGTAGCTG 1200

TTGTGCTTGT GGCCTGTGGT CAAGGAACTG CTTCTAAAGA CAACAAAGAG GCAGAACTTA 1260

AGAAGGTTGA CTTTATCCTA GACTGGACAC CAAATACCAA CCACACAGGG CTTTATGTTG 1320

CCAAGGAAAA AGGTTATTTC AAAGAAGCTG GAGTGGATGT TGATTTGAAA TTGCCACCAG 1380

AAGAAAGTTC TTCTGACTTG GTTATCAACG GAAAGGCACC ATTTGCAGTG TATTTCCAAG 1440

ACTACATGGC TAAGAAATTG GAAAAAGGAG CAGGAATCAC TGCCGTTGCA GCTATTGTTG 1500

AACACAATAC ATCAGGAATC ATCTCTCGTA AATCTGATAA TGTAAGCAGT CCAAAAGACT 1560

TGGTTGGTAA GAAATATGGG ACATGGAATG ACCCAACTGA ACTTGCTATG TTGAAAACCT 1620

TGGTAGAATC TCAAGGTGGA GACTTTGAGA AGGTTGAAAA AGTACCAAAT AACGACTCAA 1680

ACTCAATCAC ACCGATTGCC AATGGCGTCT TTGATACTGC TTGGATTTAC TACGGTTGGG 1740

ATGGTATCCT TGCTAAATCT CAAGGTGTAG ATGCTAACTT CATGTACTTG AAAGACTATG 1800

TCAAGGAGTT TGACTACTAT TCACCAGTTA TCATCGCAAA CAACGACTAT CTGAAAGATA 1860

ACAAAGAAGA AGCTCGCAAA GTCATCCAAG CCATCAAAAA AGGCTACCAA TATGCCATGG 1920

AACATCCAGA AGAAGCTGCA GATATTCTCA TCAAGAATGC ACCTGAACTC AAGGAAAAAC 1980

GTGACTTTGT CATCGAATCT CAAAAATACT TGTCAAAAGA ATACGCAAGC GACAAGGAAA 2040

AATGGGGTCA ATTTGACGCA GCTCGCTGGA ATGCTTTCTA CAAATGGGAT AAAGAAAATG 2100

GTATCCTTAA AGAAGACTTG ACAGACAAAG GCTTCACCAA CGAATTTGTG AAATAATGAC 2160

AGAAATTAGA CTAGAGCACG TCAGTTATGC CTATGGTCAG GAGAGGATTT TAGAGGATAT 2220

CAACCTACAG GTGACTTCAG GCGAAGTGGT TTCCATCCTA GGCCCAAGTG GTGTTGGAAA 2280

GACCACCCTC TTTAATCTAA TCGCTGGGAT TTTAGAAGTT CAGTCAGGGA GAATTGTCCT 2340

TGATGGTGAA GAAAATCCCA AGGGGCGCGT GAGTTATATG TTGCAAAAGG ATCTGCTCTT 2400

GGAGCACAAG ACGGTGCTTG GAAATATCAT TCTGCCCCTC TTGATTCAAA AGGTGGATAA 2460

GGCAGAAGCT ATTTCCCGAG CGGATAAAAT TCTTGCGACC TTCCAGCTGA CAGCTGTAAG 2520

AGACAAGTAT CCTCATGAAC TTAGCGGTGG GATGCGCCAG CGTGTAGCCT TACTCCGGAC 2580

CTACCTTTTT GGGCACAAGC TCTTTCTCTT AGATGAGGCC TTTAGCGCCT TGGATGAGAT 2640

GACAAAGATG GAACTCCACG CTTGGTATCT TGAGATTCAC AAGCAGTTGC AGCTAACAAC 2700

CCTGATCATC ACGCATAGTA TTGAGGAGGC CCTCAATCTC AGCGACCGTA TCTATATCTT 2760

GAAAAATCGC CCTGGGCAGA TTGTTTCAGA AATTAAACTA GATTGGTCTG AAGATGAGGA 2820

CAAGGAAGTC CAAAAGATTG CCTACAAACG TCAAATTTTG GCGGAATTAG GCTTAGATAA 2880

860

GTAGAAAAAT AGGGAGTTGG TGAAGATTAT CCTTTACCAG CGCCCTTTTT CTTTTAAAAA 2940

TGAGAAAATT TCGGTATAAT AGTCAAACAA GGTCAAGGTT TAAAGAGAGA GGTGGGTTTG 3000

TTATGAGATT TAAAAATACA TCGGATCATA TTGAGGCCTA CATCAAGGCG ATTTTAGATC 3060

AATCTGGTAT CGTGGAGTTG CAACGGAGTC AGTTGGCAGA TACCTTTCAG GTTGTTCCTA 3120

GTCAGATTAA CTACGTGATC AAGACACGCT TTACGGAAAG TAGAGGCTAC TTGGTTGAAA 3180

GTAAGCGTGG TGGCGGAGGC TACATTCGTA TAGGACGGAT TGAGTTTTCT AGTCATCATG 3240

AAATGCTCCG GGAGCTGCTT TACTCGATTG GTGAGCGAGT CAGTCAAGAA ATTTATGAGG 3300

ATATTCTCCA GCTTTTGGTT GAGCAGGAAT TGATGACCAA GCAGGAGATG AATTTGCTAG 3360

AATCAGTAGC TTTGGATCGC GTTTTAGGAG AAGAAGCTCC AGTTGTTCGA GCAAACATGC 3420

TACGTCAGAT CATACAAGAG GTAGATAGAA AAGGGAAGTA AGATGAACTA TTCAAAAGCA 3480

TTGAATGAAT GTATCGAAAG TGCCTACATG GTTGCTGGAC ATTTTGGAGC TCGTTATCTA 3540

GAGTCGTGGC ACTTGTTGAT TGCCATGTCT AATCACAGTT ATAGTGTAGC AGGGGCAACT 3600

TTAAATGATT ATCCGTATGA GATGGACCGT TTAGAAGAGG TGGCTTTGGA ACTGACTGAA 3660

ACGGACTATA GCCAGGATGA AACCTTTACG GAATTGCCGT TCTCCCGTCG TTTGCAGGTT 3720

CTTTTTGATG AAGCAGAGTA TGTAGCGTCA GTGGTCCATG CTAAGGTACT AGGGACAGAG 3780

CACGTCCTCT ATGCGATTTT GCATGATAGC AATGCCTTGG CGACTCGTAT CTTGGAGAGG 3840

GCTGGTTTTT CTTATGAAGA CAAGAAAGAT CAGGTCAAGA TTGCTGCTCT TCGTCGAAAT 3900

TTAGAAGAAC GGGCAGGCTG GACTCGTGAA GATCTCAAGG CTTTACGCCA ACGCCATCGT 3960

ACAGTAGCTG ACAAGCAAAA TTCTATGGCC AATATGATGG GCATGCCGCA GACTCCTAGT 4020

GGTGGTCTCG AGGATTATAC GCATGATTTG ACAGAGCAAG CGCGTTCTGG CAAGTTAGAA 4080

CCAGTCATCG GTCGGGACAA GGAAATCTCA CGTATGATTC AAATCTTGAG CCGGAAGACT 4140

AAGAACAACC CTGTCTTGGT TGGGGATGCT GGTGTCGGGA AAACAGCTCT GGCGCTTGGT 4200

CTTGCCCAGC GTATTGCTAG TGGTGACGTG CCTGCGGAAA TGGCTAAGAT GCGCGTGTTA 4260

GAACTTGATT TGATGAATGT CGTTGCAGGG ACACGCTTCC GTGGTGACTT TGAAGAACGC 4320

ATGAATAATA TCATCAAGGA TATTGAAGAA GATGGCCAAG TCATCCTCTT TATCGATGAA 4380

CTCCACACCA TCATGGGTTC TGGTAGCGGG ATTGATTCGA CTCTGGATGC GGCCAATATC 4440

TTGAAACCAG CCTTGGCGCG TGGAACTTTG AGAACGGTTG GTGCCACTAC TCAGGAAGAA 4500

TATCAAAAAC ATATCGAAAA AGATGCGGCA CTTTCTCGTC GTTTCGCTAA AGTGACGATT 4560

GAAGAACCAA GTGTGGCAGA TAGTATGACT ATTTTACAAG GTTTGAAGGC GACTTATGAG 4620

AAACATCACC GTGTACAAAT CACAGATGAA GCGGTTGAAA CAGCGGTTAA GATGGCTCAT 4680

CGTTATTTAA CCAGTCGTCA CTTGCCAGAC TCTGCTATCG ATCTCTTGGA TGAGGCGGCA 4740

GCAACAGTGC AAAATAAGGC AAAGCATGTA AAAGCAGACG ATTCAGATTT GAGTCCAGCT 4800

GACAAGGCCC TGATGGATGG CAAGTGGAAA CAGGCAGCCC AGCTAATCGC AAAAGAAGAG 4860

GAAGTACCTG TCTACAAAGA CTTGGTGACA GAGTCTGATA TTTTGACCAC CTTGAGTCGC 4920

TTGTCAGGAA TCCCAGTTCA AAAACTGACT CAAACGGATG CTAAGAAGTA TTTAAATCTT 4980

GAAGCAGAAC TCCATAAACG GGTTATCGGT CAAGATCAAG CTGTTTCAAG CATTAGCCGT 5040

GCCATTCGCC GCAACCAGTC AGGGATTCGC AGTCATAAGC GTCCGATTGG TTCCTTTATG 5100

TTCCTAGGGC CTACAGGTGT CGGGAAAACT GAATTAGCCA AGGCTCTGGC AGAAGTTCTT 5160

TTTGACGACG AATCAGCCCT TATCCGCTTT GATATGAGTG AGTATATGGA GAAATTTGCA 5220

GCTAGTCGTC TCAACGGAGC TCCTCCAGGC TATGTAGGAT ATGAAGAAGG TGGGGAGTTG 5280

ACAGAGAAGG TTCGCAATAA ACCCTATTCC GTTCTCCTCT TTGATGAGGT AGAGAAGGCC 5340

CACCCAGATA TCTTTAATGT TCTCTTGCAG GTTCTGGATG ACGGTGTCTT GACAGATAGC 5400

AAGGGACGCA AGGTCGATTT TTCAAATACC ATTATCATTA TGACATCGAA TCTAGGTGCG 5460

ACTGCCCTTC GTGATGATAA GACTGTTGGT TTTGGGGCTA AGGATATTCG TTTTGACCAG 5520

GAAAATATGG AAAAACGCAT GTTTGAAGAA CTGAAAAAAG CTTATAGACC GGAATTCATC 5580

AACCGTATTG ATGAGAAGGT GGTCTTCCAT AGCCTATCTA GTGATCATAT GCAGGAAGTG 5640

GTGAAGATTA TGGTCAAGCC TTTAGTGGCA AGTTTGACTG AAAAAGGCAT TGACTTGAAA 5700

TTACAAGCTT CAGCTCTGAA ATTGTTAGCA AATCAAGGAT ATGACCCAGA GATGGGAGCT 5760

CGCCCACTTC GCAGAACCCT GCAAACAGAA GTGGAGGACA AGTTGGCAGA ACTTCTTCTC 5820

AAGGGAGATT TAGTGGCAGG CAGCACACTT AAGATTGGTG TCAAAGCAGG CCAGTTAAAA 5880

TTTGATATTG CATAAAAGAA TAAAAGTATC AGCATCTGAC CATAAGTCAC AGTGGAGTGA 5940

AATTCAATGA AAATCAAAGA GCAAACTAGG CAGCTAGCCG CAGGTTGCTC AAAACACTGG 6000

TTTGAGGTTG CAGATAGAGC TGACGTGGTT TGAAGAGATT TTCGAAGAGT ATGAAACTAA 6060

AACCTATAGC TTCTAAACGA TCCGTGGTTT TCATCATTCA ACACAAAATT CATATGTTTA 6120

TTACCCTCCG TCGTATTTGT CTTAGAGCGT GTGTAGTAGA AAAAGAGCAG TCTTATCTGA 6180

AATTTTTATT CTTTCAAAAG AGACCTGTTT CTTTTTTGCA TGTCAAATCC GTTCTAGCTG 6240

GTATTTGAAA AATCAAACTA ATATTCAATG AAAATCAAAG AACAAACTAG GAAGCTAGCC 6300

GCAGGTTGCT CAAAACACTG TTTTGAGGTT GTAGATAGAG CTGACGTGGT TTGAAGAGAT 6360

TTTCGAAGAG TATAAGCTGC AAGATGAATG ATTTTCTTGT ATTGACGTTG TTGTTGACAA 6420

AAAGTAGCGG ATAAATGAAA TCCATTCCAT TATCATAGAT GATAGGCTGG TAGGAAATTT 6480

TCAAATAGCA TACAGGAAAT AGATGTATGG AGTTCTGGTA GTAGAAAGGG AGAGAGATGA 6540

ACATTTTAGT TGCAGATGAC GAGGAAATGA TTAGAGAAGG AATTGCAGCA TTTCTGACAG 6600

AAGAGGGTTA TCATGTCATT ATGGCTAAGG ATGGACAAGA GGTCTTGGAA AAATTTCAAG 6660

ATCTCCCTAT CCATCTCATG GTACTGGATT TAATGATGCC TAGGAAGAGT GGTTTTGAAG 6720

TGTTAAAAGA AATCAATCAA AAGCACGATA TTCCTGTCAT CGTCTTGAGT GCTCTGGGAG 6780

ATGAAACTAC TCAGTCACAG GTATTTGATC TCTATGCTGA TGATCATGTG ACAAAACCTT 6840

TTTCTTTGGT ACTGCTTGTC AAGCGTATTA AGGCGCTTAT CAGACGTTAC TACGTCATAG 6900

AGGATCTTTG GCGATATCAG GATGTAACAG TGGATTTTAC CTCTTACAAA GCACATTATA 6960

AAAATGAAGA AATTGATCTC AAACCAAAGG AATTACTGGT ACTAAAGTGT TTGATTCAGC 7020

ATAAAAATCA AGTTTTAAGT AGAGAGCAGA TATTGGAAGA AATTTCAAAA GATGTAGCTG 7080

ATTTACCTTG TGATAGGGTC GTTGATGTCT ATATTCGTAC TCTTCGCAAA AAATTAGCTT 7140

TAGATTGTAT CGTGACTGTG AAAAATGTTG GGTATAAGAT TAGCTTATGA TAAAAAATCC 7200

TAAATTATTA ACCAAGTCTT TTTTAAGAAG TTTTGCAATT CTAGGTGGTG TTGGTCTAGT 7260

CATTCATATA GCTATTTATT TGACCTTTCC TTTTTATTAT ATTCAACTGG AGGGGGAAAA 7320

GTTTAATGAG AGCGCAAGAG TGTTTACGGA GTATTTAAAG ACTAAGACAT CTGATGAAAT 7380

TCCAAGCTTA CTCCAGTCTT ATTCAAAGTC CTTGACCATA TCTGCTCACC TTAAAAGAGA 7440

TATTGTAGAT AAGCGGCTCC CTCTTGTGCA TGACTTGGAT ATTAAAGATG GAAAGCTATC 7500

AAATTATATC GTGATGTTAG ATATGTCTGT TAGTACAGCA GATGGTAAAC AGGTAACCGT 7560

GCAATTTGTT CACGGGGTGG ATGTCTACAA AGAAGCAAAG AATATTTTGC TTTTGTATCT 7620

CCCATATACA TTTTTGGTTA CAATTGCTTT TTCCTTTGTT TTTTCTTATT TTTATACTAA 7680

ACGCTTGCTC AATCCTCTTT TTTACATTTC AGAAGTGACT AGTAAAATGC AAGATTTGGA 7740

TGACAATATT CGTTTTGATG AAAGTAGGAA AGATGAAGTT GGTGAAGTTG GAAAACAGAT 7800

TAATGGTATG TATGAGCACT TGTTGAAGGT TATTTATGAG TTGGAAAGTC GTAATGAGCA 7860

AATTGTAAAA TTGCAAAATC AAAAGGTTTC CTTTGTCCGC GGAGCATCAC ATGAGTTGAA 7920

AACCCCTTTA GCCAGTCTTA GAATTATCCT AGAGAATATG CAGCATAATA TTGGAGATTA 7980

CAAAGATCAT CCAAAATATA TTGCAAAGAG TATAAATAAG ATTGACCAGA TGAGCCACTT 8040

ATTAGAAGAA GTACTGGAGT CTTCTAAATT CCAAGAGTGG ACAGAGTGTC GTGAGACCTT 8100

GACTGTTAAG CCAGTTTTAG TAGATATTTT ATCACGTTAT CAAGAATTAG CTCATTCAAT 8160

AGGTGTTACA ATTGAAAATC AATTGACAGA TGCTACCAGG GTCGTCATGA GTCTTAGGGC 8220

ATTGGATAAG GTTTTGACAA ACCTGATTAG TAATGCAATT AAATATTCAG ATAAAAATGG 8280

GCGTGTAATC ATATCCGAGC AAGATGGCTA TCTCTCTATC AAAAATACAT GTGCGCCTCT 8340

AAGTGACCAA GAACTAGAAC ATTTATTTGA TATATTCTAT CATTCTCAAA TCGTGACAGA 8400

TAAGGATGAA AGTTCCGGTT TGGGTCTTTA CATTGTGAAT AATATTTTAG AAAGCTATCA 8460

AATGGATTAT AGTTTTCTCC CTTATGAACA CGGTATGGAA TTTAAGATTA GCTTGTAGAC 8520

AGATTAGTTT TTTATTAAAG TTCATATAGG GTTAACATAA GTGTGTTATT CTTTGTGTAG 8580

ATAAAAGAAA GGATACTAAT ATGGTATTAG CGATTATTTT AGTAACATTC TTTATTCGAT 8640

TGATTTTTTT AAAGCGTTCG ATAGAGAATG AGAAACGAAT CCTTAGCAAT GGCGGGG 8697

(2) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4317 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

AACCATACAT ACGGCAAGGC AAAGCTGACG CGGTTTGAAG AGATTTTCGA AGAGTATTAG 60

TTGCCTTTAA AGGCATCCAC CATCGTTTGA AATTCTTCAT TTGAGAGAGT AATCCCTTTG 120

CCCATTTTAG TATGGTCTGG ACTCCAAGCA CGAATATCAA ACTTTGCAGG GGCACCATTA 180

AAGCTCACAC GGTTAATTTC CTTGGTCCAA CCTTTTTCGT TTTCAGAAAG AGTCAACAAG 240

TGCTCTTCGA TTTCAAATGT AAATTCTGCC ATTTTCTTCT CCTTTTTTAG TTTCATTAGT 300

TTATTCGTAA AATCTTGTAG ATTTTAGGAA AATTTTATAT AATATTGATA TAAAAGAAGG 360

GAGGCCAATA TGAGACATAA ATTCCAGCAA GTTCTAAATA AAATACATGA TTTTTTAAAT 420

GGATATGACC AACCTGACCA GACTGAAACC AACTCCCTTA CAGCCACTAT TGAAGAGGCT 480

ATCCAGAAAC AAACCGCTGT TCACCTTATC TTGTCTGAGA CAAGCTTTAC AGGTGACATC 540

ATCAAATATG ATCAGCAAGG CCAGCAAATT ATCGTGAAAA ATTTTTTCCAA AAATGTGAGC 600

CGGATTATCC GTATAAGCGA TATTCAACGC CTGCGATTTG TCCCCTCAAC TGTCCAAACA 660

GCCCAAAAAA ATAGATTTAA GAAAGAGTGA GATGTAGTTG CTTCATCCCA CTCTTTTTTC 720

TTAGCGAATT TGTTCAAAAT GTAAATGAAC TGCGATATGA TCTCCATAAC CACTTCTTTC 780

CAAGTCACGT TGTAAACGAT AGGAAATGTA GTGTTCTGCA ATGGTAATGT AACCTGCGCC 840

CAATAAACGA TGTTCAACCA TAGATTGAAT CATACTGATA GTCGCACGTT CCACCTTGGC 900

TTCTTGTAAA TCCAAAACTA CCTTCTTAGT GACTTGAGCA AGATTTTGAC GCAAATCATC 960

TGTCAAAACA TAAACAGTTT GGGCTGCCTT CAAGATGGCT TGGTAAATCT TATCTGGATT 1020

AAATTCAGCA ATTTCGCCAT TACGTTTGAT TACTTGCATA GGTTTCTCCT TTATTCTTTG 1080

TTTTCTTTGA TTTCTGCCAG CATTTTTTCT TCTTCTACTG TCAGTTGATA ATGTTCAAGT 1140

AAATCCGGTC TGCGCTCGTA GGTTTTCTTT AAACTCTCGT ACAATCGCCA CTGACGAATC 1200

TTTTCATGGT GGCCACTCAT CAATACATCT GGCACGACCA TGCCTCGATA ATCATAGGGA 1260

CGTGTGTACT GAGGATATTC TAAAAGACCT GAAGAAAAAC TATCATCTTG GTGGCTAGAC 1320

TCCTTGCCAA TCACTTCTGG AATCAGGCGA ACTGTAGCAT CAATCATGGT CATAGCTGCC 1380

AATTCTCCAC CAGTGAGGAC ATAGTCACCT AGGGAAATCT CATCTGTTAC CAAGGTCTTA 1440

ATGCGCTCAT CATAACCCTC ATAGTGCCCA CAGATAAAGA TTAGCTCTTC CTCTTGAGCC 1500

AAATCTTCAG CATAAGCCTG ATCAAACTGC TTTCCAGCAG GATCAAGGAG AATAACGCGC 1560

GGATTTTTCT TTTCAATAGC ATCAAAGGAA TCGAAAATAG GTTGTGCTCT GAGCAACATG 1620

CCCTGACCGC CTCCGTAGGG CTCATCATCT ACATGACGGG CCTTTTCAGC ATTTTCTCGA 1680

AAATTATGAT ACTGGATATC CAAGAGCCCT TTTTCTCGAG CCTTTCCAAC GATTGAGTGC 1740

TCCAGTGGAG AAAACATCTC TGGAAAGAGG GTTAAAATAT CAATCTTCAT CGTCTAACCC 1800

TTCTAAGATT TCCACATCGA CCCGTTTACT TGGAATATCA ACATTGAGAA CCACTGGTGG 1860

GATATAAGGT AAAAGCAAAT CACGTTTGCC TTTTCGTTTG ACCACCCAGA CATCATTAGC 1920

ACCTGGTTGC AGGATTTCCT TGATGGTTCC AACCAAGCTA TCACCCTCAT AGACTTCCAA 1980

ACCGATAATC TCGTGATAGT AAAATTCACC ATCGTCTAGG TCATTCAAAT CTTCCTCAGC 2040

GACCTTGAGA CTGTATCCCT TGTACTTTTC GATAGTATTG ATATGGTACA TATCTTTGAA 2100

TTTAATAATG TCAAAGTTCT TCTGTTTACG GTGGCTAGCG ATGGTCACTG TTTGGACAAA 2160

CTGATCTTTT TCATCAAACA AAACCAGCTC AGCTCCTTTT TTAAACCGTT CTTCTGCAAA 2220

ATCCGTCACA GACAAGACTC GCATCTCCCC CTGTAATCCC TGCGTATTAA CGATTTTCCC 2280

AACATTAAAG TAGTTCATCT TGTCTCCTGT AATCTCCTTT TTTCCATCTT ATTCTAACAA 2340

TTCTCGAATA ATAGCCGCAA TTTTTTCCGA TTCTGACCAT TGTAAATAAT GGTGATTCCC 2400

TCCTAAAATG AGTTTAGTAT TGGAAGTCCA ATATTCTGAT TCTCTGTACT CTTTTTCTCT 2460

ATAAGGCTGA CAAAAAACAA ATACAGGAAT ATGAGCTTCT ATAGATACAT CCTCAAAATC 2520

TTCCTCAGTA ATCTCTCCAG ATATCTGAAA TTCTGGATCT TGATTTTCCA ACTCTAAGCC 2580

TTTTTCTTGC ATTAATTCCC AGATTTTTTT ATTCGTTTCA GGACTAAATG TTGCTTGAGT 2640

TAAGTTCTTA AAATAAAGTT CAGGACCACA CTCGTCAATC AGCCTCATCT GCTCTTCCAT 2700

TTCTGGATAA GGATTTTCTG AAAAATCAGC AAACATGACT TTTTTAGTTG TCGGTTCAAT 2760

TGCTACTAAA GTCTGACGCT TAATTGGTTT CTCGAGTAAT TTGCAAGCTA AAATTCCACT 2820

CCAACTATGT GCACAAAGTA TATATTCAGA AATTCCTAAT TCTTCAAGTA CTTCATAAAC 2880

CGCATCTGCA AGATTATCTA GATTTTTTCC AGCTTGGTCA TGAATCGGAC TCCTACCTGT 2940

GTTCGGAAAA TCAATTGTCA AATAACCAAT TGTAGGAGGA GGTTTTTCAA GTATAAGTGA 3000

AAAATTTTCA TAACTTGGTA GCAAACCTGC TCCGTTTAAA CAAACTAGCA CTTTCTTTTG 3060

CTTTTGATAA GTAACAGAGA GGCTACCAAT TTCTGTAGAT ACTTCAAACC TCTTCATAAA 3120

GAAATCCACT GATTCTATAT AATGAATTAT TAAAAATCCT TATCCTTTAT TTTATCACGT 3180

TCCAAGGATT TTCTCAAGTT GGAGGAAGGG GACAATATCT CTACTTTCCC TTCAATAATC 3240

CTTCCAAATT ATGTTTATGT TGGTAATTAA TGGCTGCGGT TTTGTCTTTC TCAAAGACAG 3300

TCTTGGTAAG GTCAATATGA TTAATAGCTA CGATTGCGAC GGTGTAGTAA ATGATATCAG 3360

CCAGTTCTCT GGCAAGTTCC TCGTTCGAAT CCTATCCCTT CTTTTCGACC AGAGCGCCTA 3420

TTCAAAACCT CGACTACTTC TCCGACTTCC TCCACTAACT TCATAAAGAG ACCTTCATCA 3480

GTCCGAGACT GCTGTTAATG TTCGATTAAG TAGTCTTGGA ATTGCCTAAA CGTTCAATCT 3540

TTTATAGTAT ATTGAAACTA GAATAGTACA CCTTTACTTC TAAAACATTG TTAGAAATCG 3600

ATTTGACTGT CCTGATCGAT TTGTCCTGTT CTTGTTTCAT TTTACTATAT CTTCTATTCC 3660

ACACAAAAAA GCGAGACATC CGTCCCGCCC TTCTTATTTT TCGTCAATAA CGATTCTTAC 3720

TTTTTTGTAT TCAGTTGGGA CAGAGTAGAC AATCGTTCTT ATCGCAGAAA TAGTGCGACC 3780

CTTACGACCG ATTACACGAC CCACATCGCT TTGATCAAGA TTCAAATGAT ATTCCAAAAA 3840

TTCTGGTGTA TCCTCAATCT TGATAGTTAA GGCATCTGGT TGTGAAATTA AGGGTTTCAC 3900

AATCGCAATA ATGAGATTTT CAATCGTATC CATCTGTCAA CCTACTTTAA ACTTATTTTG 3960

AAAATTTAGA ATCGTGGAAT TTTTTCAATA CGCCTTCTTT TGAAAGGATG TTACGTACTG 4020

TGTCTGAAGG TTGAGCTCCA TTAGCCAACC ATGCAAGAAC GCGGTCTTCT TTCAAAGTTA 4080

CTTGGTTTTC AGCAACAAGT GGGTTGTAAG TTCCAACTGT TTCGATGAAA CGTCCGTCAC 4140

GTGGTGAACG TGAATCTGCT ACGTTGATAC GGTAGAAAGG TTTTTTCTTA GAACCCATAC 4200

GAGTCAAACG GATTTTAACT GCCATTTTTA AAGTCTCATT TCTTTAATTT TTTATTTCGG 4260

TGAAATAGCT GAGCTATTTA GCACATGTTC TATTATAGCA GATTTCTGGC ATGTGTC 4317

(2) INFORMATION FOR SEQ ID NO: 125:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4881 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

AATTTATTTG ACTGGAAATT GTAGAGGGTT CTCGAAATTT CTTGAATGGT TAAAATAAGG 60

ACAAGAGAAA ACATGGATAT CTATATCCTT GTGCCAAAAA AACCACTGCC CTCCCCAGAC 120

CAACCTGAGG AAAGCAGTGA TTCTTATTTT AGGAGTTAGG AATGAATACA CGAAATCAAT 180

TTAGCTGATT ATTTTTTGTT TTTCAAGAAT TCATCGTATT GTTTTTGCAT TTCGTTCAAT 240

ACTTTTTCGT AGGCACCTTC AGATTTCAAT TTTTCCATCA ATTCTGGAAT CGCTTTATCT 300

GGGTCTACAG TACCAGTGTT GATAGCTGTA TCAAATTGTT GCATTGTGTT AGCAATAGCT 360

GAGATTTCAG ATTTCACATT GTCAGTATTG AAGATAAATC CAAGCGCTGG AGATTCTTTA 420

GCTTCTGCCA ATTCTTTCTT AGAATTTTCG ATTTGTTGGT CTGTAACGTT TTCGTTGATG 480

TAAAGGATCC AGTTGTTACC AGTGTTCCAT CCACCCATGT GAGTGTTTCC TTTGTAGCCA 540

TCAAGAACGC GAACACGGTT TTCTTTACCT TCAATTTTTT CCCAGTTCTT GCCTTCTGGA 600

CCGTAAACAA GACCGTTCAA GAGTTCTGGG TTCGTATTCA AGAGGTTCAA GATTTCCATT 660

GATTTTTCTT TGTTCTTAGA GTTGTTTGAG ATGACAAAGT TAGCAACTTG TGTTGTTTGG 720

TTTTTCTTGA TGAAGTTAGT AATTGGTTTG ATTTGGATAT CTTTGTTGGC AACACGTGAA 780

AGCAAGCTGT TACCGTAGTC AGCTGGTCCT ACTGTTTCTT CACGAACGAA CCAAGTATCT 840

TGTTGAAGGT CAAAGGAAGT ATCGCTTGTT GCGACGTCTT TTGGAATGTA GCCAGCTTCA 900

TAGAATTTGT GAAGAGTCTT CAAGTGTTCT TTGAAACGAG GCACTTCGTA ACGGTTTACA 960

ACTTTAGTAG TATCGCCTTC AAGGTCGATA ACGAATGGAA GACCGTTTGC TACTGGGTAG 1020

TCAAAATTAT CAGATGGGAT GAAAACTTTA CCAATAGCAA ATGGTACTAC GTCTGGAGCT 1080

TTTTCTTTGA TTTGTTTCAA GACTGGCTCA AGAGTTTCGT AAGAAGTAAC ACCTGAAATA 1140

TCGATACCAT ATTTAGCAAG GAGAGTTCCG TTGAAGGCAA AGTTTTGAGA TGATGCAACG 1200

TTGGCTGCAA CTGGAACAGC GTAAATCTTA CCATTTACAG TATTACCCTT GATGTAAGCT 1260

GGGTCAAGTG CTTTGTAAAG GTCTTTACCT TCTTTTTTGT ACAATTCTGT CAAGTCAGCG 1320

TAAGCACCTT TTTGAGCATT TACAATATAG TTATCTGCAA AGGCAATATC ATAGTTTTCA 1380

CCAGATGATG TGATAACTGA CATTTTCTTA CCATAGTCAC CCCAGCCAAG GTATTGGATA 1440

TCCAATTTGG CACCAACTTT TTCTTCAATG ATTTTGTTGG CATTTGCTAA CAATTCATCC 1500

AAGTTGTCTG GTTTGTCACC GATTTGGTAC ATTTTGATAA CAGGTTTGTC ACCTGAATCA 1560

GCAGCTTTTT TGCTGTTACC TGTCAAATTT CCACAAGCAG CAAGACCTGC AGCCAGAGCG 1620

ACTACACTAG CAGATGCAAA AGCATATTTT TTCCAGTTTT TCATGATAAA AACTCCTTTT 1680

TTTATTTTTA AACTTATAAA CAATGTAATG ATCTTATACT CAATAAAAAT CAAAGAGCAA 1740

ACTAGAAAAC TAGCCGCAGG CTGCTCAAAG CACTGCTTTG AGGTTGTAGA TAAGACTGAC 1800

GAAGTCAGTT ACATATATCT ACGGCAAGGC GACGTTGACG CGGTTTGAAT TTGATTTTCG 1860

AAGAGTATTA ACTTCACACA AGGGAAGTTG GGAACTGAGA AATGTTATTT CTCAATAAGC 1920

ACTATTCTTT CACACCACCG ATAGTCAAAC CTTTTACAAA GTAGCGTTGG AAAAATGGAT 1980

ACAAAATCGC GATTGGAAGG GTTGCAACCA CAACCATGGC CATACGACCT GTTTCTTTCG 2040

GTAGAGCAAC TCCCAGTTGA CCAATCAAGC CGACCGCTTT GGCAATGTAG TCCATATTTT 2100

GTTGGATTTG CATGAGCAAA TATTGCAATG GATACAAGTT GTCACTCTTG ATGTAAAGAA 2160

GGGCGTTGAA CCAGTCATTC CAGAAACCAA GAGCTGTTAA GAGCGTGATG GTTGCGATAC 2220

CTGGTAGTGA CAATGGCAAA CAGATTTGGA AGAAAATCCG GGCCTCACTG GCACCATCGA 2280

TACGAGCCGA TTCTAGAATG GCTTCTGGAA TGGTCTTCTT GAAGAAGGAA CGCATCAAGA 2340

TGATGTTAAA TGGTGAGAGA AGCATTGGAA CAATCAAGGC CCAAACAGTG TCACCAAGCT 2400

GAAGTACACG GGTCACCATG ATATAACCTG GTACCAAACC AGCGTTGAAC AACATACTGA 2460

GAAGGACGAA GATGGTAAAG AATCTGCGAT ACTTAAAGGT TGTCCGTGAA ATAGCGTAGG 2520

CATAGGTTGT TGTGATAAAG ACATTTGTCA ATGTCCCAAC TACGGTTACA AAGACAGAGA 2580

TGAAGAGGGC TTGTAGGATT TTATCCTTAA ACTGTGCCAA AAACTCAAAA CCGTCTAAGC 2640

CAAATTGGGA TGGGAAGAAG CTATAGCCGT ATTGGAGGAG GCTTTTCTCG TCTGTCACTG 2700

AAATAATGAT AACGAATACA AAAGGTAGGA TACAAGAGAG GGCAATCAAA CCCGAAATGA 2760

TACTGAAGAA GATATCTGCT TTCTTACTGA AGGAGTGAAT GCCGACATTA TCAATTTTTT 2820

CTTTTTTAAT TTTCTTTTTT GCCATATTCT CCTCCTTTCT AGAACAAAGC TGAGTTTGGA 2880

TCGACTCGTC TTGCAAGCAA GTTTGATAGG ATAACCAGAA TCAAACCAAC AACGGATTGG 2940

TAAAGACCGG CTGCTGCAGC CATACCGATA TCTGCTGTCT GAGTCAAACC ATTAAAGACA 3000

TATACGTCCA AAACGTTGGT TACATTGTAA AGCTGACCAG CATTGTGTGG GATTTGATAG 3060

AAGAGACCGA AGTCTGCGCG GAAGATATTT CCGACTGCAA GGATGGTCAA TACAGTTACA 3120

AGCGGAGTCA ACTGAGGAAT GGTTACGTTG CGAATACGTT GCCACTTGCT AGCTCCGTCC 3180

ACTGTCGCTG CTTCGTAGTA GGTTGGATCA ATTCCCATGA TCGTCGCATA GTACATGACA 3240

CTGCTATATC CAAAGCCTTT CCAAATACCT AGGAAAAGTA GGAGATAGGG CCAGATGCCC 3300

AGGTCAGCGT AGAAATTGAC TTCTTTGAGA CCAAGACTTT CCAATAGATG ATTGAACACC 3360

CCTTTATCAA TATTTAGGAA GGCATCTGTA AAGAAACTGA TGATAACCCA AGACAAGAAG 3420

TAAGGGAACA ACATAGAAGT TTGAAAAATC TTCACCATTC TCTTAGAACG GAGCTCGCTG 3480

AGGATAATGG CAATCCCTAC AGATACAACT AAACCTAGAA AGATAAAGCC AAGATTGTAG 3540

AGGACAGTAT TTCGTGTGAT AATAAAGGCG TCTCTTGAAC TAAATAAGAA TCTAAAATTA 3600

TCGAGTCCGA CCCATTTACT ATTTATGATA CTATCTATGA AACCATTACT GGTCATGTGG 3660

TAGTCTTTGA AGGCAACCAC GTTCCCAAAT ACTGGAATGT AAAAGAATAG AATCAACCAG 3720

AGTGCCCCTG GCAAAACCAT CAAGAGAAAG ATCCAGTTGT CTCTCAATGT TTTTGAAAAC 3780

TTTTTCATAA TTTCCTCCCT TTTTATTTTG ATATCCATCT AAAAATTCTT TTTTAGACTT 3840

TTGATAACGA TTACATTATT AGTATACTCC TATTTGCAGG TTAGGTTAAA CTCCTAATTA 3900

TAGAAAAAAC TCCACAAATT ATGTAGCAGA TTTAAAACTT TATCACCACT ATCAAACAAA 3960

TGTCCTAAAT CAATTGTTTA TTTTATCTCT ATTAGCCCAG TGATGGCGTC ACTCTGTTAT 4020

AAGCATCCAA CAACGGGGTA TACTGAAAAA TCTCCAGACT AGGGAACTCA GCGATAGTTC 4080

CTAATCTGGA GATTTTTAAT ATGTTATTAG GCGTTTGCTT TCAACTTAGC AATAACCTCT 4140

TTAAGATTAT CAATCAACTC TGCTGCAGTA TGCTCAGAGC CTTTTTCATC TGCCAAGAAC 4200

AAAACTGCTT TTTGAAGTTC TTTTTGAGAG TTTTCAAGGA CATCCTTATC TACTGTTTCA 4260

AGGTTTGAGT CTTTAAGAAG TTTACTTAAT TCCTTGGCTA ATTTCTTGAG TTTGATTTGC 4320

AGACTCATCT TCTCCTGCTG TTTCTTTGCC CGCTGTTTGT CCTCCATCCT TAGTTGCTGA 4380

CTGGCTTTCC TTAATGGACT CTAGGGAAGC AATGGCATCT TTGACTGTTT GCAAGATATC 4440

ACGTAAACCT TGCTCTGTCA AACTATCATC TGCAAAAGCT TTATTAGCCT CTGCCAAAAC 4500

CAGACGTGCT GAATCTGTGG TAGGATTCGA TACACCTGTC AATGATCTCA AAAGATTTTC 4560

TAAGGTTTGA GTCTGCTTAC TAATACTAGA CTAAAATCAA AAAGTATTAT ATAACAGTGA 4620

TATGAAATCA ACTAAAGAAG AAATCCAAAC CATCAAAACA CTTTTAAAAG ACTCTCGTAC 4680

AGCTAAATAT CATAAACGCC TTCAAATCGT TCTATTTTGT CTGATGGGCA AATCTTATAA 4740

AGAGATTATA GAACTTTTAT AGTAGTTTGA AATAAGATGT GAACATCTCT ATCAGGAAAG 4800

TCAAATTAAT TTATAGAAAT ATTTTAGCAG CCAAGGTGTA CTGTTATAGA TTCAATACAC 4860

TATACTTGGT GGTTTAGCTC G 4881

(2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13121 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

AGGATCCCCG GAAAAGGAGA CTAAAAATGA AGAAAAAATT TCTAGCATTT TTGCTAATTT 60

TATTCCCAAT TTTCTCATTA GGTATTGCCA AAGCAGAAAC GATTAAGATT GTTTCTGATA 120

CCGCCTATGC ACCTTTTGAG TTTAAAGATT CAGATCAAAC TTATAAAGGA ATTGATGTTG 180

ACATTATTAA CAAAGTCGCT GAGATTAAAG GCTGGAACAT TCAGATGTCC TATCCTGGAT 240

TTGACGCAGC AGTCAATGCG GTTCAAGCTG GGCAAGCCGA CGCTATCATG GCAGGGATGA 300

CAAAGACTAA AGAACGTGAA AAAGTCTTCA CCATGTCTGA TACTTACTAT GATACAAAAG 360

TTGTCATTGC TACTACAAAG TCACACAAAA TTAGCAAGTA CGACCAATTA ACTGGCAAAA 420

CCGTTGGTGT TAAAAACGGA ACTGCCGCTC AACGTTTCCT TGAAACAATC AAAGATAAAT 480

ACGGCTTTAC TATTAAAACA TTTGACACTG GTGATTTAAT GAACAACAGC TTGAGTGCTG 540

GTGCCATCGA TGCCATGATG GATGACAAAC CTGTTATCGA ATATGCCATT AACCAAGGTC 600

AAGACCTCCA TATTGAAATG GATGGTGAAG CTGTAGGAAG TTTTGCTTTC GGTGTGAAAA 660

AAGGAAGTAA ATACGAGCAC CTGGTTACTG AATTTAACCA AGCCTTGTCT GAAATGAAAA 720

AAGATGGTAG TCTTGATAAA ATTATCAAGA AATGGACTGC TTCATCATCT TCAGCAGTGC 780

CAACTACAAC TACTCTCGCA GGATTAAAAG CTATTCCTGT TAAGGCTAAA TATATCATTG 840

CCAGCGATTC TTCTTTTGCC CCTTTTGTTT TCCAAAATTC AAGCAACCAA TACACTGGTA 900

TTGATATGGA ATTGATTAAG GCAATCGCTA AAGACCAAGG TTTTGAAATT GAAATCACCA 960

ACCCTGGTTT TGATGCTGCT ATCAGTGCTG TCCAAGCTGG TCAAGCCGAT GGTATCATCG 1020

CTGGTATGTC TGTCACAGAT GCTCGTAAGG CAACTTTTGA CTTCTCAGAA TCATACTACA 1080

CTGCTAATAC CATTCTTGGT GTCAAAGAAT CAAGCAATAT TGCTTCTTAT GAAGATCTAA 1140

AAGGAAAGAC AGTCGGTGTT AAAAACGGAA CTGCTTCTCA AACCTTCCTA ACAGAAAATC 1200

AAAGCAAATA CGGCTACAAA ATCAAAACCT TTGCTGATGG TTCTTCAATG TATGACAGTT 1260

TAAACACTGG TGCCATTGAT GCCGTTATGG ATGATGAACC TGTTCTCAAA TATTCTATCA 1320

GCCAAGGTCA AAAATTGAAA ACTCCAATCT CTGGAACTCC AATCGGTGAA ACAGCCTTTG 1380

CCGTTAAAAA AGGAGCAAAT CCAGAACTGA TTGAAATGTT CAACAACGGA CTTGCAAACC 1440

TTAAAGCAAA CGGTGAATTC CAAAAGATTC TTGACAAATA CCTAGCTAGC GAATCTTCAA 1500

CTGCTTCAAC AAGTACTGTT GACGAAACAA CGCTCTGGGG CTTGCTTCAA AACAACTACA 1560

AACAACTCCT TAGCGGTCTT GGTATCACTC TTGCTCTAGC TCTTATCTCA TTTGCTATTG 1620

CCATTGTCAT CGGAATTATC TTCGGTATGT TTAGCGTTAG CCCATACAAA TCTCTTCGCG 1680

TCATCTCTGA GATTTTCGTT GACGTTATTC GTGGTATTCC ATTGATGATT CTTGCAGCCT 1740

TCATCTTCTG GGGAATTCCA AACTTCATCG AGTCTATCAC AGGCCAACAA AGCCCAATTA 1800

ACGACTTTGT AGCTGGAACC ATTGCCCTCT CACTCAATGC GGCTGCTTAT ATCGCTGAAA 1860

TCGTTCGTGG TGGTATTCAG GCCGTTCCAG TTGGCCAAAT GGAAGCCAGC CGAAGCTTGG 1920

GTATCTCTTA TGGAAAAACC ATGCGTAAGA TTATCTTGCC ACAAGCAACT AAATTGATGT 1980

TGCCAAACTT TGTCAACCAA TTCGTTATCG CTCTTAAAGA TACAACTATC GTATCTGCTA 2040

TCGGTTTGGT TGAACTCTTC CAAACTGGTA AGATTATCAT TGCTCGTAAC TACCAAAGTT 2100

TCAAGATGTA TGCAATCCTT GCTATCTTCT ATCTTGTAAT TATCACACTT TTGACTAGAC 2160

TAGCGAAACG CTTAGAAAAG AGGATTCGTT AATGGCAAAA TTAAAAATTG ATGTAAATGA 2220

TTTACACAAG CACTATGGAA AAAATGAAGT CCTAAAAGGA ATTACGACTA AGTTCTATGA 2280

AGGAGATGTT GTTTGTATCA TCGGTCCTTC AGGTTCTGGT AAGTCAACTT TCCTCCGTAG 2340

CCTCAATCTT TTAGAAGAAG TCACTAGCGG TCACATCACT GTGAACGGCT ATGATTTAAC 2400

TGAAAAAACA ACCAATGTTG ACCACGTCCG TGAAAATATC GGCATGGTAT TCCAACACTT 2460

CAACCTCTTC CCTCATATGT CTGTATTGGA CAACATCACC TTTGCTCCTA TTGAGCACAA 2520

GTTGATGACT AAGGAAGAAG CTGAGGAATT GGGAATGGAG TTGCTTGAAA AGGTTGGACT 2580

AGCAGATAAA GCTAATGCCA ATCCAGATAG CCTATCAGGT GGTCAAAAAC AACGTGTGGC 2640

CATCGCTCGT GGCCTAGCAA TGAATCCAGA CATCATGCTC TTCGATGAAC CAACTTCTGC 2700

CCTTGACCCT GAGATGGTTG GAGACGTACT TAACGTTATG AAGGAATTGG CTGAGCAAGG 2760

CATGACCATG ATTATCGTAA CCCATGAGAT GGGATTTGCT CGTCAGGTTG CCAACCGCGT 2820

TATCTTTACT GCAGATGGCG AGTTCCTTGA AGACGGAACA CCTGACCAAA TCTTTGATAA 2880

CCCACAACAC CCTCGTCTGA AAGAGTTCTT AGATAAGGTC TTAAACGTCT AAACTCAAAC 2940

TGTAAGGATT TCCTTGCAGT TTTTCTACCT CGTATTGGAA TTTTTGATTT TTCGGAAAAT 3000

TATGTTAGAA TTAAGTTTAT GAAATGAGGT TTCCTCATAC CTAGCAAGAC TAGGAATAAA 3060

AATAGAAATT AGGTAGCTAG ATGTCATCTA AGGTTATTGT TACAATTTTC GGTGCGAGTG 3120

GAGACCTGGC TAAACGCAAG CTCTACCCTT CCCTTTTTAG ACTATATCAA TCCGGCAATC 3180

TTTCCAAGCA CTTTGCCGTT ATTGGAACTG CCCGTAGACC TTGGAGTAAG GAATATTTTG 3240

AATCTGTAGT TGTCGAGTCC ATCCTTGATT TGGCAGATAG TACCGAGCAA GCCCAAGAAT 3300

TTGCTAGCCA CTTCTACTAT CAAAGCCATG ATGTCAATGA TTCGGAACAT TATATTGCTT 3360

TGCGTCAATT ACAAGCTGAG CTTAATGAAA AATACCAAGC TGAACACAAT AAGCTCTTCT 3420

TCTTGTCTAT GGCACCTCAG TTCTTTGGAA CCATTGCCAA ACACCTCAAA TCTGAAAACA 3480

TTGTCGATGG CAAAGGTTTT GAGCGCTTGA TCGTTGAAAA ACCATTTGGT ACAGATTACG 3540

CAACTGCAAG CAAGTTGAAT GACGAACTCC TAGCAACATT TGACGAAGAA CAAATTTTCC 3600

GTATCGACCA TTATCTTGGT AAGGAAATGA TCCAAAGCAT CTTTGCAGTT CGCTTTGCAA 3660

ACTTGATTTT TGAAAACGTT TGGAACAAGG ATTTTATCGA CAATGTTCAA ATTACCTTTG 3720

CGGAGCGCTT GGGTGTAGAA GAACGTGGTG GCTACTATGA CCAATCCGGT GCCCTCCGTG 3780

ACATGGTCCA AAACCACACT CTACAACTTC TTTCGCTCCT CGCCATGGAC AAACCAGCAA 3840

GCTTCACAAA AGACGAGATT CGTGCTGAAA AGATTAAGGT CTTTAAAAAC CTCTATCATC 3900

CAACTGATGA AGAACTCAAA GAACACTTTA TCCGTGGGCA ATACCGCTCT GGTAAGATTG 3960

ATGGCATGAA ATACATCTCT TATCGTAGCG AGCCAAATGT GAATCCAGAA TCAACAACTG 4020

AAACCTTTAC ATCTGGTGCC TTCTTTGTAG ACAGCGATCG ATTCCGTGGT GTTCCTTTCT 4080

TTTTCCGTAC AGGTAAACGA CTGACTGAAA AAGGAACTCA TGTCAACATC GTCTTTAAAC 4140

AAATGGATTC TATCTTTGGA GAACCACTTG CTCCAAATAT TTTGACCATC TATATTCAAC 4200

CAACAGAAGG CTTCTCTCTT AGCCTAAATG GGAAGCAAGT AGGAGAAGAA TTTAACTTGG 4260

CTCCTAACTC ACTTGATTAC CGTACAGATG CGACTGCAAC TGGTGCTTCT CCAGAACCAT 4320

ACGAAAAATT GATTTATGAT GTCCTAAATA ACAACTCAAC TAACTTTAGC CACTGGGATG 4380

AAGTTTGTGC GTCATGGAAG TTGATTGACC GTATTGAAAA GCTCTGGGCT GAAAATGGTG 4440

CCCCACTTCA TGACTATAAA GCTGGAAGCA TGGGACCTCA AGCCAGCTTT GACCTACTTG 4500

AAAAATTCGG TGCCAAATGG ACTTGGCAAC CAGATATCAC CTATCGTCAA GATGGTCGCT 4560

TAGAATAAAA AAATTTCCTG CAAGTTTATG CcTTGCAGGA TTTTTGCTTC TGATTAGATT 4620

AAACCTTCCA AGAGACCTTT CATAAAGTTT TCTGAGTTAA ACTCTCCAAT ATCATCGATT 4680

TTTTCACCAA AACCAATCAA TTTTACAGGA ATATTGAGTT CTTCACGAAT GGCTAGAACC 4740

ACACCTCCTC GAGCAGTTCC ATCAATCTTA GTCAAAACAA TTCCCGTTAA AGGTGTGATT 4800

TTCGAAAATT CTTTGGCCTG TACTAGGGCA TTTTGACCTG TTGATGCATC AAGTGCCAAG 4860

AAGGTTTCAT GTGGTGCTTC TGGCACAACA CGTTTGATAA TACGACCAAT CTTTTCCAAC 4920

TCAGCCATAA GGTTATCCTT ATTTTGCAGA CGACCAGCAG TATCAATCAT GAGAATATCG 4980

ATACCTTCAG TCACGGCACG TTCCATACCA TCAAAGACCA CGCTGGCTGG ATCAGCTTTT 5040

TCAGGTCCAG TTACTACTGG AACATCTACT CGTCGGCCCC ATTCAGCTAG CTGAGCTACT 5100

GCACCCGCAC GGAAGGTATC TGCTGCAACC AGCATGACCT TCTTACCAGC TTGTTTGTAG 5160

CGGTGGGCTA GTTTTCCGAT AGAAGTTGTT TTCCCAACAC CATTCACACC AACAAAGAGC 5220

ATAACTGTCA AGTTATCTTG GAAGTGGATG CTTTCATCGT AGCTACCATC CTTTTCATAA 5280

AGCTCAACCA ATTTCTCAAT GATGACACGA CGAAGTACAT CAGGTTTCTT GGCATTTTCA 5340

AGCTTGGCTT CGTAACGTAG TTCCTCCGTT AAGTTAGAAG CGACTTGGAC ACCAACATCA 5400

CTCATAATCA GCAGTTCTTC CAGTTCCTCG AAAAATTCTT CGTCAACAGA GCGGAAGTTA 5460

GCAAAGAAGG CATTCAAGCG GGCACCGAAA CCTGTGCGAG TTTTCTTAAG ACTGCGGTCA 5520

TATTTTTCCT GAACAGTTTC TTCTGTTTGA GGAGCTTCTG GTTCAAGCAC TTCAGAATTA 5580

TTTTCTTCTA CAGTTCCTTC GTGCTCAAGC TTCTCTTCCT CTGGTAATTC TTCTGAGTTT 5640

GGTAATTCTT CTATTTCTTC TTGAGAAACC CCTACAGCTG GCTCTGAATC CTGACTTTCT 5700

TCAACTGTGT CTTGGATTTC CTCTTCTTGG AACACAGCTT GTTCAACAAT TTCAACCTCT 5760

GCTTCTTCCT GAGAAACTTC CTCAACTTCT GTGAAGGTAG GATCAACATC TTCAGACAAA 5820

TCAAGATTTT CCAGAGCTTC TTTTACAACT TCTTCGATTT TAGGTTCTTC TTTTTTTCCG 5880

AATAGACGGT CAAACAATCC CATATCTTAG TTCTCCTTTA GCACATATTC TTCGATAGCC 5940

CAGGCGACAG CTTCCTCATC GTTGGTCATC GGCGTCACTA CATTTGCGGC TGCCTTTACT 6000

TCAGGAACAG CGTTTTGCAT AGCAACACCA AGACCTGCCC ATTCAATCAT AGAGAGGTCA 6060

TTGGCCTCGT CACCACAAGC CATCACTTGA CTTTGGTCGA TTCCAAGATG GCTGATTAGT 6120

TTTGCCAAAC CTGTTGCTTT ATGAACATTC TTTGGTGACC ATTCTAGCAA CATTTCACGT 6180

GATTTAAAGA TTTCATATTG GTCAAACAAT TCTGGAGAAA TCTTCTGAAT GGCTGCATCC 6240

AAGGGTTCTT GAGCAAAGGC AGTCACGCAT TTGTTGTAGG TCATTTGACT AGATAAGTCT 6300

TCAAAGTCCA CTGGAACAAA GGTCAAAGCT GGATTGAATT TGGCATAAAG ACTTTCTTGG 6360

TCCGATTGGA TTTGATAAAC TGTTCCTTCT GAGATGGCAT CAAGAGGCAG TGATAATTTC 6420

TCTGTTTCTT CATACAAACG TGCCACATCA TCATATGAAA AGACTGTTTT ATCAAGGATT 6480

TCTCCTGTAT TTTTCTGAAC TAATCCACCA TTAAAAGTAA TGGTATACTC ATCTTCCTGA 6540

CCGTCAGTCC CTAACTCATG GAGAAAGAAA TCCATGGCTT TTAAGGGACG ACCAGTTGTC 6600

AATACGACCT TGATACCACG ATCACGCGCA gCTTGCAAGG TTTCCTTGGT ACGATCCGTC 6660

AGCCTTTTAT CAGTAGTCAG CAAGGTCCCG TCCAAGTCCA ATGCAATCAA TTTTATATCT 6720

GCCATTATAA GCCCTCCATA TAAGCTATAA CCGACCGTTC CTTATGGTGA CCAATCACAG 6780

TCTTTGCTAA TTCTAAAATT TCAGGTCGTG CATTTTCAGG AGCTACAGGA TGTCCCACAA 6840

CCTGCATCAT ATGTAAGTCA TTAAGATTGT CTCCAAAAGC CATGACCTGA TCCATTGTGA 6900

TACCAAGTTT TTTAACTAAT TCAACAATGG CCACTCCCTT ATCGACATAG TCCAGAACAA 6960

TATCAATGGA TTCAAAGCCA GTTGTCATGG CCTTAACACC AGGAACGTTT TCGTTTACCC 7020

AAGCCTCCCC ATCTTCCAGC GTTTCTTCTG TGAAGTTGGT TGTAAATTTG AAAATGTCAT 7080

CTGTGATATC TTCCAAACTC GCTACTTTTT GGATATTTTC ATTATAGTGC TGACTCACTT 7140

TCAAATAGGT CTCATCAACC GTATCTAGAA CATATGAACC CTTCTTACCC GTCAAGAGCA 7200

GTTTATTGAT ATCTACATAA GGTGAAGTTT TCAGCTTTTC AAAAGTTGCC AGATAAAAGT 7260

CACGAGACAT AGTCGCTTCA TACAAGTCCT GACCTTGATA CTCTACCAAA CTGCCATTTT 7320

CCGCGATGAA AATAATGTCA TCACGAACAC CAGCAAATAA TTTTTCTAGA GACAGAAATC 7380

CCCGACCCGA AGCTACCGCA AAGTAAATCC CTTTTTCCTT GTAGGAAACC AAGAGAGACT 7440

TGAGACGATC CATATCAAAG CGTCCATTCC CATCTAGGAA GGTTCCGTCC ATATCCGTTG 7500

CTACTAGTTT AATTGTCATC CTTCAATACT TTCTAAATCT TTTAACTTAA CTGAAACAAT 7560

CTTTGAAACA CCCGATTCTT GCATGGTCAC TCCATAGATG GAATCAGCCG CTGCCATGGT 7620

TCCCTTACGG TGGGTTACGA CGATGAACTG GCTGTCCTTG TCAAAGCGGT TGAGGTAATC 7680

CCCAAAACGT TTAACATTGG CTTCATCCAG CGCAGCTTCC ACCTCATCCA AGATAACAAA 7740

TGGAATAGTC TTGACACGAA TAATGGAGAA GAGCAAGGCA AGAGCCGATA GGGCTTTTTC 7800

ACCACCACTC ATGAGATTAA GAGACTGGAT TTTCTTGCCT GGTGGTTGGA CAGAAATTTC 7860

AACCCCAGCT GTCAGCAAGT CTCCTTCAGT CAAAATGAGG TCAGCCTGAC CTCCACCAAA 7920

CATCTGCTTG AAGGTCACTT TAAAGGACTC ACGAATGACC TCAAAGGTTG ATTTAAAGCG 7980

TTCCTTGACC TCATCATTCA TCTCTGTAAT GGTCTCAAGG AGCAGGTTTT TCGCAGACAA 8040

AATATCATCA CGTTGGCTAT TTAGGAAATC CAGACGGTTG TGAACTTCTT CGTACTGTTC 8100

AATAGCGTCT AAATTGACAG GACCCAGTGA GCGTATAGCC TTCTCTAAAT CCTTAACTTC 8160

TTGCTCTGCC AGATTGAGAT TTTCCAACTC ATGCGCCTTT TCTAAAGCTT CTGTGTAGCT 8220

GATCTGGTAC TGGTCTGTTA ATTGACTTTG TAGATGGCGC AAGCGCTCGC TAACCTTTTC 8280

TTTCTTGGCT TCAGCACGAG TTTGCTTGCG AATCCACTCT TCATTCTGCT GGCGAGCCTG 8340

ATCCAAATGA CTAGCAATAT CATCCAGTTG ACCCTCAATA TCATCCAACT CAAACTGCTT 8400

GCGAATCAAA CCTTGTTGGA GATTTGTTTT TTGAGTTTTG GATTCTTCCG CCTGTTGACT 8460

GAGCAATTCT GTATCAACCT TCTCAAGATT ATCAATCTTT TCTTGAAGAA GGCGCTGGAT 8520

TTCCTCTTGT TCAAAATCAA GATTGTCCAA TTCCTTGCCT AAGCGTTCAA TATCAGCAAC 8580

TTCATAACGT TTTTGCCCTT GCAGTTCTGT CTTAAGCAAA CGAGCTTGCG CTAGCTCTTC 8640

CTGCAAGTTT TGATAGCGTT CTTGGATGGC ATTTTTGTTA GACTTAATCT CTTCAATCTC 8700

AGCTTCCAGA TTTTGCTTGT CACTGGAGAT TGCAGCAAGA CGCTCTTGGC AGTTTTCCTT 8760

ATCCGCTTGC CAATCTCCCT CGGAAAGACG ATCTATTTCC TCTTCTTGGA GTTTCCAAAG 8820

AGTTTCCAGT TCTTCAACTT GCTGACTAGT TTGCTGATAA GCGAGGAACA AGCCTTGCTC 8880

CTGAATACGT GCCTGCTCTC CTTGAGATTT AATAGCTTCT AATGACTCGG TCAATCTGGC 8940

CATCTCATCT TGCAAGGTCT TCAAAGTCGC CTCTTCTGAA CCCAAGCTTG CTTCTTCTTC 9000

AGCAATTTCT TTTTGTAATT GCTCCAGTTC TGGCTTGATA AAAATGCTGT TATTCTGGCG 9060

ATTGGCACCA CCTGCATAAG AACCACCTGT GCGCAACTCT GTCCCATCCA ATGTCACCAT 9120

ACGAACCTGA TAACGAACTT GGCGAGCTGC TGCACGCGCA TGTTCTACGG TATCAAACAT 9180

AGCCGTCGTA GCTAGCAAGT TCTTGAAAAT GGCTTCCAGT CTAGTATCAA AAGTCACCAA 9240

CTCATCTGCC ATCCCAAGGA AACCTGGGCT TACAGCGATA GCATCTTGGT TCTGACTAGA 9300

AATCGTACGC GCCTTGATAG TGGTCAAAGG AAGAAAGGTT GCACGACCGG CTCTGTTCCG 9360

TTTAAGGAAG TCAATAGCCT TGGTTGCCGA CTCTTCATCT TCTACGATGA TATGCTGGCT 9420

ACTTGCCCCT AAGGCAATCT CTAGGGCAGT TTGATAATAA ACATCAAAGG TCAGATGCTC 9480

ACTGACTGCA CCAATAATCC CACCTAGGCG ATCTTTTTCT TGGAGAACAC TCTTAACACC 9540

TGCATAAAAG TTACTATGAT TTCTCAGGAT ATTTTCCAAA CTTTGAGCTC TGGCCTGCTT 9600

GTTTTTGAGA TTATCCAGAC GGTCAAAGAG TTGGCTTTGT TGAGCTTGAT AGGAAGTTTT 9660

CTGCTCCTCT TGCTCCTTGG CAATAGCTTG GTAGTCAGCC AATAATTTCT GAACCTGCTC 9720

CTTGGCAGTT TCAAGCTCTT CCTTTTGCTG ACTAGCCTTC TCTTTAGCTA TAGCTAATTG 9780

CTCTTTCAGC TTTTCTAGTT GATCTGCTTG TTTTTGAGAA AGCTGACGAC TATTTTCCAA 9840

CTCATTCTCA ATACGGGTCA ACTGGTTTGA GACATCCGCT TCTTCTTGTA AAAGAGCTAC 9900

AAAGCGTTCA CGTAAGAGCT CAATCATCTG ATCAGGATCG TCTGAGAAAG CCAGCAATTC 9960

AGCTTCTAAA CGATTGAGTT TTTGATTATT TTGGACTAGA TTTCCCTCTA ACAGAGCTAA 10020

AGAGCTTTCT TTATCAGACT TTTCTTTGCT GAGTGAATTT CTCTTATCCT CCAAAGCAGC 10080

CAAACGGGCT TGTGCCTCCT GTTGATTCAA GGCCACTTGC TCGGACTCCA GTTTCGATAG 10140

GGCTAATTTT CTTTCTAAAT CACTAATCAG ACTAGTCAAG TCCATCAAAC TGCCTTGGTC 10200

TTTGGCCATT TCAGCCTGTA AATCTTGGCG TTGCTTTTTA AGAGTTTGAT TTTCTTCTTC 10260

TAATTTTTCA CGCTTTTGGT AATAACTCAT CAAGAGTTCT TGAACCTGAG TCAACTCTTC 10320

TTCTGTCGAC TCTAGTTCAG CCTTATTTTC CTTGATTTGA GCAACCAGAA CATCTAAATA 10380

AATAGCCTTA CGTTGTCCTT CCAAGTCTAA AAACTTACGG GCATTCTCAG CTTGCTTCTC 10440

AAGAGGCTTG ATTTGATTAT CCAACTCGTA GATAATGTCC TCTAAGCGGT CCAGATTATC 10500

CTGAGTTTGC TGCAGTTTAC TCTCGGTTTC TTTTCTGCGA GTCTTGTATT TTAAAACTCC 10560

AGCAGCTTCT TCAAAAATAG CTCGTCGTTC CTCAGGCTTG GAATTAAAAA TCTCCTCAAC 10620

CTTCCCTTGG GAAATAATAG AGAAGGAATC TCGTCCCAAT CCAGTATCCA AGAAGAGGTC 10680

ATGAATATCA CGCAGACGGA CTTTCTTGCC GTCAATCTTG TATTCGCTAT CTCCACTACG 10740

ATAGACATGG CGTTCCACCC TGATTTCTTG ACCTGCATCC TTGATAAATC CGTCATGATT 10800

ATCCAGAGTC ACAACTACAG AAGCATAATT GAGCGGTTTG CGACTTTCGG TTCCAGCAAA 10860

GATGATATCC GGCATCTTGC CCCCACGGAG ACTCTTGACA CTAGACTCCC CCAAAGCCCA 10920

ACGCAGACTT TCTGTAATAT TGGACTTTCC AGATCCATTG GGTCCAACAA CTGCCGTCAC 10980

ACCTTGGTCA AAAACGACCT TGGTCTTATC AGCAAAAGAC TTGAACCCCT GAATTTCGAT 11040

TTCCTTTAAA TACATGAATC CAGCCCCTTC TCAACGGCAT TTTTGGCAGC TTCCTGCTCT 11100

GCTAATTTCT TAGAACGACC TTGGCCTTGA CCGATGCTCT TACCTTCAAC AAGAACTTCT 11160

ACATCAAAAA CCTTATCGTG AGCAGGCCCT GTTTCAGAAA TCACCTGATA ACGAATAGCC 11220

ACATCACCAT TGACCTGAAG CAACTCTTGG AGATGGGTTT TATAGTCTGT AATCATCTCA 11280

AACTCGCCTG CTTCAACCTT AGGAATCATG ACTTGATAGA TAAATTCCTT GACCTTGGCC 11340

ACATCCTTAT CCAAAAGAAG GGCACCAAGA AAGGCTTCAA AGGCATCACC AAGAATGGTG 11400

TCACGATTGC GACCACCTGA TTTTTCTTCC CCTTTACCCA ACTTGATAAA CTGGTCAAAC 11460

TGGCAATCAC GCGCAAAACC AGCTAAACTC TCCTCACGGA CAATCATAGC ACGGAGTTTT 11520

GATAGGTCAC CTTCAGGCTT TTTAGGATAT TTTTTATATA GATATTCTGA AATCAATAAC 11580

TGTAGAACAG CGTCTCCTAA AAATTCCAAG CGTTCATTGT GTGAAATTTT TAAGAGGCGG 11640

TGCTCATTGG CATAACTCGT ATGAGTAAAG GCAGTTTCCA GTAACTTTTT GTCTGCAAAT 11700

TCGATTGCAA AATGATTCTT TAGTACAGTT TGTAATTCTT TCATACCAAC CTCTTTCTAA 11760

CTGATAATAG TCCTTTTTAT TATATCAAAA AAAGCCCCCT GAGTCACTCT AAAACGGGAC 11820

TGGAAAGCAT TTGGGAATTC TTTAGACAGA GATTCTCAGT TTTAGCGGCA AATTTGGGTC 11880

AGGATAAAGA AAAAAGCCCT ATTAAAGGCT TTTTAGGATG TTTACATCCA CCCTGAGGGA 11940

ATCGAACCCC CATCTCAAGA ACCGGAATCT TACGTGATAT CCATTACACT AAGGGTGGAA 12000

ACTTGTTTTA TTATAACAGA AATTTGCTCT AATAACAAGT TTTTTGGTCA AAGACCCCGT 12060

CTTAGTGGGA AGCATCCCCA TTCCAGATGG AGTTTTTCAC GATCACATAA TCAACGTGTT 12120

TAAGGTCAGC AACCTGACGT CCACCTGCAT AAGAAATAGC ACTTTGAAGG TCTTGTTCCA 12180

TCTCAGTTAA AGTGTCTTGC AGATGACCTT TAGCAGGAAG CAAGATACGT TTGCCTTCCA 12240

CATTTTTGTA AGCACCTTTT TGATATTGTG AGGCTGAACC ATAATATTCT TTGAACTGTT 12300

CACCATCGAC TTCAATCGTT TTCCCTGGAC TTTCAATGTG TCCTGCAAAG AGGGAACCAA 12360

TCATGATCAT GCTAGCACCG AAGCGGATAG ACTTAGCAAT ATCACCGTGA GTACGAATTC 12420

CTCCATCAGC GATAATCGGT TTACGCGCAG CCTTGGCACA CCAGCGTAGA GCAGCCAACT 12480

GCCAACCACC TGTACCAAAA CCAGTCTTAA CCTTGGTGAT ACAAACCTTA CCAGGACCGA 12540

TTCCGACCTT AGTAGCATCC GCACCAGCAT TTTCCAATTC ACGCACAGCT TCTGGTGTTC 12600

CCACATTTCC AGCAATGACA AAGGTATCTG GCAATTCTTT CTTGATGTGT TGAATCATAG 12660

AAATCACGCT ATCCGCATGA CCATGAGCAA TATCAATAGT GATATACTCA GGAGTATCAG 12720

CCTTGAGCTG GCTAACAAAA TCATACTCAT AATCCTTAAC ACCGACAGAG ATAGAAGCAA 12780

TGAGCCCTTG ATTGTGCATT CGTTTAATAA AAGGAATGCG TCCTGCCTCA TCAAAACGGT 12840

GCATAATGTA GAAGTAACCA CCTTTAGCCA GTTGCTCTGC TACATTTTCA TCCAAAATCG 12900

TCTGCATATT CGCTGGCACA ACAGGTAGTT TAAAGGTGTG ATTTCCTAAA GTGACACTTG 12960

TATCCGCTTC TGCACGGCTT TTAATGACAC ATTTATTTGG AATCAATTGA ATATCTTCGT 13020

AATCAAAAAT TGGAAATTCA TTTAACATAT CGATGTCTCG TTTCTTTTGT AATGACCTAC 13080

CTATGCTCTT GCATCACTAC GCCTTTTCCG ACGTTTCCTG G 13121

(2) INFORMATION FOR SEQ ID NO: 127:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9578 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

CCGAATGCAA TGTTTACGGT TGAACTTGAA AATGGACATC AGATTTTAGC AACAGTTTCT 60

GGTAAAATTC GTAAAAACTA TATTCGTATT TTAGCGGGAG ATCGTGTTAC TGTCGAAATG 120

AGTCCATATG ACTTGACACG TGGACGTATC ACTTACCGCT TTAAATAATC GAAAAACTTG 180

GAGGGATAAG AAATGAAAGT AAGACCATCG GTCAAACCAA TTTGCGAATA CTGTAAAGTT 240

ATTCGTCGTA ATGGTCGTGT TATGGTAATT TGCCCAGCAA ATCCAAAAACA CAAACAACGT 300

CAAGGATAAG ATAGAAAGGA GAAAACATGG CTCGTATTGC TGGAGTTGAT ATTCCAAATG 360

ACAAACGCGT AGTAATCTCA TTGACTTATG TTTATGGTAT CGGACTTGCA ACATCTAAGA 420

AAATTTTGGC TGCTGCTGGA ATCTCAGAAG ATGTTCGTGT ACGTGATCTT ACATCAGATC 480

AAGAAGATGC TATCCGTCGT GAAGTGGATG CAATCAAAGT TGAAGGTGAC CTTCGTCGTG 540

AAGTAAACTT GAACATCAAA CGTTTGATGG AAATCGGTTC ATACCGTGGT ATCCGTCACC 600

GTCGTGGACT TCCTGTCCGT GGACAAAACA CTAAAAACAA CGCCCGCACT CGTAAAGGTA 660

AAGCTGTTGC GATTGCTGGT AAGAAAAAAT AATATAGGAG GTAAAAGTCT TGGCTAAACC 720

AACACGTAAA CGTCGTGTGA AAAAGAATAT CGAATCTGGT ATTGCTCATA TTCACGCTAC 780

ATTTAATAAC ACTATTGTTA TGATTACTGA TGTGCATGGT AATGCAATTG CTTGGTCATC 840

AGCTGGTGCT CTTGGTTTCA AAGGTTCTCG TAAATCTACA CCATTCGCTG CTCAAATGGC 900

TTCTGAAGCT GCTGCTAAAT CTGCACAAGA ACACGGTCTT AAATCAGTTG AAGTTACTGT 960

AAAAGGTCCA GGTTCTGGTC GTGAGTCAGC TATTCGTGCG CTTGCTGCCG CTGGTCTTGA 1020

AGTAACAGCA ATTCGTGATG TGACTCCAGT GCCACACAAT GGTGCTCGTC CTCCAAAACG 1080

TCGCCGTGTA TAATCATCGC ATTACACTGC TTTTCGTTTA AGAGGGAGTA ACTAAATGAT 1140

CGAGTTTGAA AAACCAAATA TAACAAAAAT TGATGAAAAT AAAGATTATG GCAAGTTTGT 1200

AATCGAACCA CTTGAACGTG GCTACGGTAC AACTCTTGGT AACTCTCTTC GTCGTGTACT 1260

TCTAGCTTCT CTACCAGGAG CAGCTGTGAC ATCTATCAAC ATTGATGGTG TGTTACATGA 1320

GTTTGACACA GTTCCAGGTG TTCGTGAAGA CGTGATGCAA ATCATTCTGA ACATTAAAGG 1380

AATTGCAGTG AAATCGTACG TTGAAGACGA AAAAATCATC GAACTGGATG TTGAAGGTCC 1440

TGCTGAAGTA ACAGCTGGTG ACATTTTGAC AGATAGCGAT ATTGAAATTG TAAATCCAGA 1500

TCATTATCTC TTTACAATCG GTGAAGGTTC TTCTCTAAAA GCGACTATGA CTGTTAACAG 1560

TGGTCGTGGA TATGTACCTG CTGATGAAAA TAAAAAGGAT AATGCACCAG TTGGAACACT 1620

TGCTGTAGAT TCTATTTATA CACCAGTTAC AAAAGTCAAC TATCAAGTGG AACCTGCTCG 1680

TGTAGGTAGC AATGATGGTT TCGACAAATT AACCCTTGAA ATCTTGACAA ATGGAACAAT 1740

TATTCCAGAA GATGCTTTAG GGCTTTCAGC ACGTATTTTG ACAGAACATC TTGATTTGTT 1800

TACAAATCTT ACTGAGATTG CTAAGTCAAC TGAAGTGATG AAAGAAGCTG ATACTGAATC 1860

TGACGACCGT ATTTTAGATC GTACGATTGA GGAACTGGAC TTGTCTGTGC GTTCATACAA 1920

CTGTTTAAAA CGTGCCGGTA TCAATACTGT GCATGATTTG ACAGAAAAAT CTGAAGCAGA 1980

GATGATGAAA GTACGAAATC TTGGACGCAA GAGTTTGGAA GAAGTGAAAC TCAAACTCAT 2040

TGATTTGGGT CTTGGATTAA AAGATAAATA AAGGAGGAAT ACATGGCTTA CCGTAAACTA 2100

GGACGCACTA GCTCACAACG TAAAGCAATG CTTCGCGATT TGACAACTGA CCTTTTGATC 2160

AACGAATCAA TCGTGACAAC TGAAGCTCGT GCTAAAGAAA TCCGTAAAAC TGTTGAAAAA 2220

ATGATTACTC TAGGTAAACG TGGTGATTTG CATGCACGTC GTCAAGCAGC TGCTTTCGTA 2280

CGTAATGAAA TCGCATCTGA AAACTATGAT GAAGCAACTG ATAAGTACAC TTCTACTACA 2340

GCACTTCAAA AATTGTTCTC AGAAATCGCA CCTCGTTATG CTGAACGTAA CGGTGGATAC 2400

ACTCGTATCC TTAAAACTGA ATCACGTCGT GGTGATGCAG CGCCAATGGC GATCATCGAA 2460

TTAGTATAAA ATCATCAATT TTGTTGAGTG TTATGATGAT GGAGTCTTGT GCTCTTAGTC 2520

TAGCTCTGGT CTACCGCTAG GATTTCGGTC CTAGCGGGAA CACTCATCAT AAGTTGGGAT 2580

AGTAGACGCT TGTTTACGAA ATTGTTTTTT TCTTAAGAAC AACTTCGTAA GCAGGCGTTT 2640

TTGAGTATTT TCGTTAGAAT TATGCTATAC TATTTGAAAA GAATCCTGTT TAATGTTAAG 2700

GTTTCTTATT TTAAGAAGAA TTGGAGTTTA CTTATGAAAG CCATTATAAC TGTTGTTGGT 2760

AAAGATAAAT CTGGAATTGT TGCAGGTGTT TCTGGTAAAA TTGCAGAATT AGGATTGAAT 2820

ATTGACGATA TCTCTCAAAC TGTCTTGGAT GAATATTTTA CGATGATGGC TGTTGTATCT 2880

AGTGATGAAA AGCAAGATTT TACCTATCTT CGTAATGAAT TTGAAGCTTT TGGGCAAACT 2940

TTGAATGTAA AAATCAATAT TCAGAGTGCA GCGATTTTCG AAGCTATGTA TAATATCTAG 3000

GAGGTCATCA TGGATATTAG ACAAGTTACT GAAACCATCG CCATGATTGA GGAGCAAAAC 3060

TTCGATATTA GAACCATTAC CATGGGGATT TCTCTTTTGG ACTGTATCGA TCCAGATATC 3120

AATCGTGCTG CGGAGAAAAT CTATCAAAAA ATTACGACAA AGGCGGCTAA TTTAGTAGCT 3180

GTTGGTGATG AAATTGCGGC TGAGTTGGGA ATTCCTATCG TTAATAAGCG TGTATCGGTG 3240

ACACCTATTT CTCTGATTGG GGCAGCGACA GATGCGACGG ACTACGTGGT TCTGGCAAAA 3300

GCGCTTGATA AGGCTGCGAA AGAGATTGGT GTGGACTTTA TTGGTGGTTT TTCTGCCTTA 3360

GTACAAAAAG GTTATCAAAA GGGAGATGAG ATTCTCATCA ATTCCATTCC TCGCGCTTTG 3420

GCTGAGACGG ATAAGGTCTG CTCGTCAGTC AATATCGGCT CAACCAAGTC TGGTATTAAT 3480

ATGACGGCTG TGGCAGATAT GGGACGAATT ATCAAGGAAA CAGCAAATCT TTCAGATATG 3540

GGAGTGGCCA AGTTGGTTGT ATTCGCTAAT GCTGTTGAGG ACAATCCATT TATGGCGGGT 3600

GCCTTTCATG GTGTTGGGGA AGCAGATGTT ATCATCAATG TCGGAGTTTC TGGTCCTGGT 3660

GTTGTGAAAC GTGCTTTGGA AAAAGTTCGT GGACAGAGCT TTGATGTAGT AGCCGAAACA 3720

GTTAAGAAAA CTGCCTTTAA AATCACTCGT ATCGGTCAAT TGGTTGGTCA AATGGCCAGT 3780

GAGAGACTGG GTGTGGAGTT TGGTATTGTG GACTTGAGTT TGGCACCAAC CCCTGCGGTT 3840

GGAGACTCTG TGGCACGTGT CCTTGAGGAA ATGGGGCTAG AAACAGTTGG CACGCATGGA 3900

ACGACGGCTG CCTTGGCCCT CTTGAACGAC CAAGTTAAAA AGGGTGGAGT GATGGCCTGC 3960

AACCAAGTCG GTGGTTTATC TGGTGCCTTT ATCCCTGTTT CTGAGGATGA AGGAATGATT 4020

GCTGCAGTGC AAAATGGCTC TCTTAATTTA GAAAAACTAG AAGCTATGAC GGCTATCTGT 4080

TCTGTTGGAT TGGATATGAT TGCCATCCCA GAAGATACGC CTGCTGAAAC TATTGCGGCT 4140

ATGATTGCGG ATGAAGCAGC AATCGGTGTT ATCAACATGA AAACAACAGC TGTTCGTATC 4200

ATTCCCAAAG GAAAAGAAGG CGATATGATT GAGTTTGGTG GTCTATTAGG AACTGCACCC 4260

GTTATGAAGG TTAATGGGGC TTCGTCTGTC GACTTCATCT CTCGCGGTGG ACAAATCCCA 4320

GCACCAATTC ATAGTTTTAA AAATTAAGAA AATAGGAGAA ATTTTAAGTT CTATTTAAGA 4380

TTAGACGTGT ATACTATAAT CATTAAATAA AGACCTCCTA ATATTATTTG AAACAGATAA 4440

CACTGAATTA GTTTGAATTT GATTTTCATC TAATATCTTT ATTTAATGAA CTCCTAAACT 4500

TTTTCATAAT AATCTCCTTC AAAAGTCGCC TGTATGGGTG GCTTTTATTT TATCATTCAT 4560

GATATAATAG AAGCAAACGG AGGACGGAAA ATGGTAAAAG TACGATTGTA TTTGGTACGT 4620

CATGGCAAGA CCATGTTTAA CACGATTGGT CGCGCGCAAG GTTGGAGCGA TACTCCCTTA 4680

ACTGCTGAAG GTGAACGAGG GATTCAAGAG TTAGGAATCG GTTTGCGAGA ATCTGATCTA 4740

CAGTTTGAGC GTGCTTATTC GAGTGATTCT GGTCGTACCA TTCAGACCAT GGGAATTATC 4800

CTTGAAGAAC TTGGCTTGCA GGGGGAAATC CCTTATCGCA TGGACAAGCG TATCAGAGAA 4860

TGGTGTTTCG GTAGTTTTGA TGGAGCCTAT GATGGCGATC TTTTCATGGG CATTATTCCT 4920

CGTATCTTTA ATGTGGACCA CGTTCACCAA TTGTCTTATG CTGAACTGGC TGAGGGCTTG 4980

GTAGAGGTCG ATACAGCTGG TTGGGCTGAA GGCTGGGAAA AACTCAGTGG CCGAATCAAG 5040

GAAGGCTTTG AAATGATTGC AAAAGAAATG GAAGATCAAG GTGGAGGTAA CGCCCTTGTT 5100

GTCAGCCATG GAATGACTAT TGGAACCATT GTTTATCTGA TTAATGGCAT GCATCCGCAT 5160

GGTCTGGATA ATGGTAGCGT GACAATCCTT GAATATGAGG ACGGCCAGTT TAGGGTTGAA 5220

GTTGTCGGTG ACCGTAGTTA CCGAGAGCTA GGACGTGAGA AGATGGAAGA AGGCTCTATT 5280

TAATCAGTCT AGACTTGCTT GCCATGAGCT AGGGATTTGA TAAGAATATC AAGATAAGAA 5340

AAAACAGCCG AGGGCACTCC TTTCGGCTGT TTTTGATGTG GAAAACTAAA GTGTAATGCT 5400

ATTGCTTTTA GAGATTTTCA TAAACAAGAG CAAGGAACCT ACTGTTAGAA CAGTCAGGAT 5460

AGTTGACAAG GTTGCGGCTA CACCGTAATT TCCTCTGAGA ACCTCTGTAT AAATAGCTAC 5520

AGTCATTGTT CTTGTTTTGA CATTGTAGAG GAGGATAGAA GTAGAGAGTT TTGAAATCAT 5580

TGTGACTCAA GATAAGATGG CTCCAGAAAT GATACCAGAT AGCATCATTG GAGTTGTAAT 5640

CTTAGCAAAG GTATTGAGAC GACTACTTCC TAAGCTTTCA GCAGCTTCTT CAATACTTGG 5700

TGCTATTTGT TGTAAGCTAG CAACAGATGA GCGAATAGTA TAAGGTAATC TTCTGGCAGA 5760

880

TAGAGACATA ATCAAGATGA AAGCAGTCCC TGTAATCATA AGAAATCCAC TTCCAAATAG 5820

ACCAGTATTG AAGGAAGAAA TGAAGGCAAT CCCTAGAACG GTTCCTGGTA CAATATAAGG 5880

TACCATACTG AGGCTGTCAA TTAAGTTTGT AAACAAATTC CGTTTTCTAA CGGCTAGGTA 5940

GGAGATAAAT GTCGCAAATA GAACAACTAG AACTAAGGCA ATCAAAGGGA TACGAATGGT 6000

ATTGAAAATA GCAGATCCCA TACGATGGAA AGCTACCTTG TAACTGTTTG GAGAATAACC 6060

TTTAACAGAT ACCATACCTG ATGTTTTTAG GAAAGAGGTA TAAATTAAGT AGATTTGAGG 6120

TAAAACAGAG ATAAAGATAA TTCCGTAGAC TGTTGCATAA ATGGCAGCCA TTTTTCCTTT 6180

TGTAGTTTTT TTAGGCTCAA TTGGATGGAG CAGATTCATG CTGAAACTGT AGCGGTTTGC 6240

AATGTGTTTT TGGATAAGGA AAATTGCCAA GGCAATGATA ATCGCCATAA TTGCAAAAGC 6300

AGAATTTCCT CCAACCTCGC TAATAAATTG GGTATAAATC AGGACAGGGA AAGTCCGATA 6360

CCCTTCGCCA ATCAACATAG GCGTTCCAAA GTCTGAGAAT GCTCTCATAA ATACAAGCAA 6420

GGAGCTGCTA GTAAGGTTGG AACTAGGAGA GGTAAAACAA CCGTTACGAT AGGTTTAAAT 6480

CCGAAGGACC CCATGCTTTC AGCTGCTTCA AGTAGAGAAT TGTCAATACT GTTCATTGTT 6540

CCAGCAACAT ATAGAAATAC CAGTGGGAAT AGTTGCAGTG TAAAGACAAG TACAATTCCT 6600

TTGAATCAAT AAATATCGAT AGCTGGAAGA TAAAGGGCAT TTGTCAAAAA TTTAGTGATG 6660

ACCTCATTTC GTCCTAGCAA GAGAACCCAG GAGTAGGCTC CTACGAAAGG AGCTGACATG 6720

GAAGCAATGA TAATCAATAT TTGTAGAAAT TTCTTCCCCT TGAAGTCATA CATAGAGAAG 6780

AGATAAGCTA ATAGGGTTCC TACAACTAAG GAAGTGATAG TAGCGGTAAT GGAAACCTTG 6840

AAACTGTTGA CTAGTGTCTC AGAGTAGTAG GCTTTACTAA AGAAAGTGAC AAAATTAGCT 6900

AGTGAGAATT GTCCTTCATG TATAAGTGCT TGCTTGAGCA CGGTAACGAT AGGATAAACG 6960

AGAAAGATAG GATAGGTAAG AAAGAGGAAG AAAGAGGAAA CTGTCCAAAT ATTTAGTTTT 7020

TTACGTTCCA TGGTTGACTC CTTTTATCAG GTTTTGGGAA CCATCTGCAG AAAAGATGTT 7080

TAATTTTTGC GTATTGATTC GTAGACGAAT ACGATTGCCT TTTTGTAGAT CTTCTTCAAA 7140

AGTTGATTCT TCACTAACTT GAATTTTTGA GGCAAAACCT GTCTCAATGA AATAATCCGT 7200

ATTTAGTCCA AGATAGACGC TATCTCTAAT AGTTCCTTCA ATATCTCCAG ATTCATCTTT 7260

GATAAACTCT TCGGGACGAA TGCTTACATG AATAGCTTGC TCCTCAACCT GATCAAGAGC 7320

TGGCATTCGA AGGGCATAGC CATCTGAAAA GACGATATAA GCGCCGTCGC TCCGTTTTTC 7380

AAGATTGGCA GGGATAATAT TTGTGCGTCC GATAAAGGTT GCCACAAACT CATTAGCTGG 7440

TTTATGATAG AGTTCTTTTG GTCGGCCGAT TTGTTGGATC ACCCCATCTT TCATAACAGC 7500

AATTTGGTCT GAAATAGCCA TGGCTTCTTC TTGGTCGTGG GTTACATAAA CAGTTGTAAT 7560

TCCCACTTCG TGTTGGATTT CTCGGATGGC TTGACGCATA TCCAAGCGAA GTTTGGCCTC 7620

CAGATTACTA AGTGGCTCGT CCATGAGGAG AACACTTGGA TTAACCGCTA AGGCGCATGC 7680

CAAGGTGACA CGTTGTTGTT GTCCACCACT GAGTTTATCG GGCTTTCGAT CCGCATATTG 7740

AGCAATTTGC ATGAGTTCAA GATACTTGTT GGTCTGTTGA ATCAATTCTT CTTTTGGAAC 7800

CTTCTTTTGC ATAAGACCAA AAGCAACGTT GTCTCGGACA GTCAAATGTG GGAAAATAGC 7860

GTAGTTTTGG AAAACCATCC CGATATTGCG TTTGCTGGGT TCCATATTAT TGATTTTTGT 7920

ATCATCGAAG TAAAATTCTC CACCTTCGAT ACTGTTGAAA CCTGCAATCA TACGAAGAAG 7980

GGTCGTTTTC CCACATCCTG AAGCTCCAAG AAGGGTAAAG AGACTTCCTT TTGGAATTGT 8040

AATGTTCAAA TTCTCAATAA CAGGGACATC GTGGTAGATT TTTTTGGCGT TAATAATTTT 8100

GATCTCACTC ATAGTGAACC TCTTTTACTG TTTAGATTGG ATATCTGTAA AGACTTCGTT 8160

GTATTTCTTA ACGATATCTG ATTTATTCTT GATGACATAA TCATAATCTT CAGTGAGTGT 8220

TTTGATTTTG TCAATTGGTT TCATGTTTTC GCTTGTTTTA GCATTTTTAC GAACAGGACG 8280

GTTAGTAGTG GTTGTACCAA GTGTATCTTG TACTTCTTGA GAGATAATAA AATCGATAAA 8340

TTTCTTGGCA TTTTCCATAT TTTTAGATTT TTTAACGATA GCAGCACTAG CAGGTAGGAA 8400

GACGGTTCCT TCTTTTGGAT AGACTACCTT AATGTTAGCT CCGTCATTTA AGAGTTTAAC 8460

TGCTGGATCT TCATAAGAGA GACCAACAGC CATTTCTCCA TCAGCGACTA CTTTATAGAC 8520

ACTAGATGAA CTTGAACCGA TTTTACCATC AATAAGTGTG AAAAGATCTT TTACATAAGA 8580

CCAAGCCTTA TCATCTTTGT AACCACCTTG AGCTTGTAGC ATATTTGTTA ATTGAGCAAA 8640

GGCGCTAGAA GAGTTTGCTG GGTCAGCAGT TGCGATTTTT CCTTTTAGTT CAGGTTTGAA 8700

AAGATCGTTA TATCCTTCGA TGTTCATGCC TTTAGTTAAA TCAGGGTTGA CGATTAAAAC 8760

ACTACCATCT AGTGTATAAG GAGTAGAGTA GCCAGTTGTG TTTTGATATT CTTTGATAAC 8820

ATTATCATTT TCTTTTGAAG TATAGTTTTC AAAGAGTTCT CCGTGGGTAG TATATTGTGT 8880

ATAAGAACCA CCAAAGATAA CATCAGCTAC AGGAACTTCT TTTTCTGACT CTAGTTTTTT 8940

GAAAAGTTCT CCAGTACCAG CTTGAATCAG TTCTACTTTG ATACCATATT TTTCTTCAAA 9000

GGCAGGAATA GTTGCTCCAA TTAAGCCCTC TGAGTTTGGT GAATAAACGA CTAGCGAACC 9060

GCCGTCTCCT TTATCAGATG AACTGTCATC GGCAGATTCA TTAGAAGAAC AAGCAGCATA 9120

ATACATCCAT TTCTTTTTCA TGATGGATAC CTCCGTTGTG TTATTTAAGT TTATTTTAAA 9180

ACAATGTAAG CGTTTTTAAA ACATACAATT CTATTCTATA GTGTATTGAA TCTATAACAG 9240

TACACTTTGA CTGCTAAAAT ATTTCTATAA ATTAATTTGA CTTTCCTGAT AGAGATGTTC 9300

ACATCTTATT TCAATTCACT ATATTAGAGT AAAATTCTCT ACAAAAAGAA GAATAGCCTA 9360

TTTTACTATT CTTCTGAGTG ATTTCAATTC CTTTGGGGAA ATATGGAGAT ACTTTTTAAA 9420

TCCTGACAAA TGGTTGTTTC TTTTTCTAAA TCGGTGATAC TGTATCGGAG AATGCGCGTG 9480

AGGTCACAAA GGCTGCGATA GAGCTTCTAT GGAGAATTTC TTTTTGGAGA GATTTTTTAA 9540

AGGAATGAGA CATCCGCTAC CTCCTTGGAA GGTTTTTG 9578

(2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13440 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

CGGGCTGTTG TGACGATTCT TATTTCTATC TGTGTTATCT TTTTGGGAAC TATTTTGGGT 60

GTTGTCTTGG CTTTTGGGCA ACGTTCAAAG TTTAAACCGC TTGTTTGGTT GGCCAACTTG 120

TACGTTTGGA TTTTCCGTGG GACACCGATG ATGGTTCAAA TTATGATTGC CTTTGCTCTT 180

ATGCATATCA ATGCTCCGAC TATTCAGATT GGAATTTTAG GTGTTGATTT TTCGCGTCTG 240

ATTCCAGGGA TTTTGATTAT CTCTATGAAT AGTGGTGCTT ATGTTTCGGA GACTGTTCGT 300

GCCGGAATCA ATGCGGTTCC AAAAGGTCAG CTAGAAGCGG CTTATTCGCT AGGGATTCGT 360

CCTAAAAATG CGATGCGTTA TGTGATTTTG CCACAAGCAG TCAAAAATAT CTTGCCAGCA 420

TTGGGGAACG AATTTATCAC CATTATCAAG GACAGCTCCC TCTTATCAGC TATTGGGGTC 480

ATGGAGTTGT GGAATGGGGC TACAACAGTT TCTACAACAA CCTATCTACC TTTAACACCA 540

CTTTTATTTG CAGCATTTTA CTACTTGATT ATGACCTCTA TTCTGACAGT AGCCTTGAAA 600

GCTTTTGAAA AACATATGGG ACAAGGAGAT AAGAAATAAT GACAGAAACC TTGATAAAAA 660

TTGAAAATTT ACATAAATCC TTTGGAAAGA ATGAAGTATT GAAGGGCATC AACCTCGAGA 720

TTAAAAGAGG AGAAGTTGTC GTTATCATCG GTCCTTCAGG GAGCGGGAAA TCTACCTTGC 780

TTCGCTCTAT GAATTTGTTG GAAGAAGCAA CCAAGGGGAA GGTTATCTTT GAGGGAGTCG 840

ATATTACGGA CAAGAAGAAT GACCTGTTTG CCATGCGTGA GAAGATGGGC ATGGTTTTTC 900

AACAATTCAA TCTCTTTCCT AATATGACTG TGATGGAAAA TATCACCTTG TCCCCTATCA 960

AGACCAAAGG TGACAGTAAG GCCGTTGCAG AGAAAAGAGC TCAGGAACTT TTGGAAAAAG 1020

TTGGTTTGCC AGATAAGGCA GACGCTTATC CACAGAGTTT GTCAGGTGGC CAGCAACAGC 1080

GGATTGCCAT CGCGCGTGGG TTGGCTATGG AACCAGATGT TTTGCTCTTT GACGAGCCAA 1140

CTTCAGCCCT AGATCCTGAG ATGGTTGGAG AAGTTCTGGC TGTTATGCAA GATCTAGCCA 1200

AGTCAGGAAT GACCATGGTT ATCGTAACAC ATGAGATGGG ATTTGCCCGT GAGGTGGCAG 1260

ATCGTGTCAT CTTTATGGCA GACGGTGTGG TTGTTGAAGA CGGAACACCT GAGCAGATTT 1320

TTGAACAAAC CCAAGGACAA AGGACTAAAG ACTTCTTGAG TAAGGTTTTA TAAGTTAGCT 1380

TTGTTTAGCT ATTTGTAGCC AGCTTTAAAC GTTAAAGAGA AGATTAGTGA AAAGCTCAAC 1440

CAGAGCTTTT TCTTATAGTT TAAAGCTATA GGATTGCCTA GGAAAGAAGT GTTAGAGCTA 1500

CATTGTATTT TTTGGTATAA TTAAAGATAT TTGTAAGAAA AGAGAAGTGA TATGACACAG 1560

ATTATTGATG GGAAAGCTTT AGCGGCCAAA TTGCAGGGGC AGTTGGCTGA AAAGACTGCA 1620

AAATTAAAGG AAGAAACAGG TCTAGTGCCT GGTTTGGTAG TGATTTTGGT TGGGGACAAT 1680

CCAGCCAGCC AAGTCTACGT TCGCAACAAG GAGAGGTCAG CCCTTGCGGC TGGTTTCCGT 1740

AGCGAAGTAG TACGGGTTCC AGAGACCATT ACTCAAGAGG AATTGTTAGA CCTGATTGCT 1800

AAATACAATC AGGATCCAGC TTGGCATGGG ATTTTGGTTC AGTTGCCATT ACCAAAACAC 1860

ATTGATGAAG AGGCGGTTCT ATTGGCTATT GACCCAGAAA AGGATGTGGA TGGTTTCCAT 1920

CCTCTAAACA TGGGGCGTCT TTGGTCTGGT CATCCAGTCA TGATTCCTTC GACACCGGCA 1980

GGAATTATGG AAATGTTCCA TGAATATGGG ATTGACTTGG AAGGTAAAAA TGCAGTCGTC 2040

ATCGGTCGAT CCAATATTGT CGGAAAACCT ATGGCCCAGC TTCTTTTGGC AAAGAATGCA 2100

ACAGTAACCT TGACTCACTC ACGTACTCAT AATCTTTCCA AGGTGGCTGC AAAAGCAGAT 2160

ATTCTGGTTG TTGCAATCGG TCGTGCCAAG TTTGTGACTG CTGACTTTGT CAAACCAGGT 2220

GCGGTAGTCA TTGACGTTGG GATGAACCGC GATGAAAATG GTAAGCTCTG TGGGGATGTT 2280

GATTATGAGG CGGTTGCCCC ACTTGCTAGC CACATTACGC CAGTCCCTGG AGGTGTCGGT 2340

CCTATGACCA TTACTATGCT GATGGAGCAA ACCTATCAGG CAGCACTTAG GACATTGGAT 2400

AGAAAATAAG ATAAAAATTT TCTGAGGAAA GTGTATTTTC TATAGCTATA TCTAAAATGA 2460

TAGAAATGAA TATTAAATTT TAGAAATAAG TTTATAAAAG GAGGTTTGCG CCTCCTTTTT 2520

GTTGTATAAT GGAGTGAGGT GATTAGATGA TTTTAAAAAT TTATAATGGG GAATATAGTT 2580

TACAATGGGA TGGAATATAC TACTTAGCAC TAATTGATTA TCCAAATATT CAAGAGTGGG 2640

AATTAGAAAA AATTGCTAAA TTTATAGCTT ACGAAAAACT TCATAAACGT CAAACAAGTA 2700

TTGAGTGTGC TGATTCTTGT TTAAAAAAAG AAATTTTAGA TTACATCTGT CAGCATCCCT 2760

TTCTGCCACC ATTTACTCCT ACAGATAAAA GAGTAGCCTC GACTTATGAC CTACATAAGA 2820

GGTTAGTGAC TTCAGACTAC TGTAGTCATA CTACGACTAT AGATGCAGCG ATTTCTATTT 2880

TTAAAACTGG TCGTCTTTTA TCTGCTGTGA AAGCCTTTGG GCGAGATGCT GAGGAGTTGG 2940

TTTTGGATAG TCGAAATGCT GCATCTGATC CGATAGATTA TTTTGACTAT GTCATGTTAG 3000

GGTGGTCAAA TACAAGTTCT GGTTATCGAT TGGCGATGGA GCGTTTATTA GGTCGAGCTC 3060

CTTCAGAGAA AGAATTACAA GACAAGTTTA TTCCTGGAGT AAGTTTTCAT TTTATCTATA 3120

CAGATTTGAT TAAAGTTCCT GGTTATATTT TTGATGGTTA CCATGCTGTA AAAATTAAGG 3180

ACATGCTTAA TTTATTAAGT GAGTTGTATA TTTGCATTAT TCCAACTCAT AATAAGAGCC 3240

AATTTGAAAA TATTATTCCA ACCAAAATAC AAGATAGGGT GTATTATCTT GACTATGCTG 3300

GAGAAGACTT AGAAGAGTGG ACTAAGAAAG TCTATCAAGT TGTTTTAAAA CAATCAGATA 3360

AAGGATAGTT GAGGAAAAAA CGATGAAAGT GATTGATCAA ACCTTACTAG AAAAAGTCAT 3420

TATTGAACGT TCTTGTACAA GTCATAAAGG AGACTACGGT CGTCTGCTGT TGCTTGGTGG 3480

GACTTATCCT TATGGTGGTG CCATCATCAT GGCTGCTTTA GCAGCTGTAA AAAGCGGTGC 3540

AGGATTGGTA ACCGTTGGAA CGGACAGGGA AAATATCCCT GCTCTACACA GCCATTTGCC 3600

TGAGGCTATG GCCTTTTCTC TGCAAGATCA GTAATTGTTA CAAGAGCAAT TGGAGAAGGC 3660

AGAAGTTGTC TTGCTGGGGC CTGGTTTACG AGACGATACG TTTGGAGAAA ATCTTGTAAA 3720

ACAGGTCTTT GCTAGCTTAA AAAAGAATCA GATTTTGATT GTAGATGGAG GGGCCTTAAC 3780

CATCCTTGCT AGGACAAGTT TGTTGTTTCC ATCTAACCAG CTTATCTTAA CTCCCCACCA 3840

AAAAGAATGG GAAAAACTGT CTGGTATTGC TATTGAAAAG CAAAACGAAG GTACAACATC 3900

TAGTGCCCTG ACTTCTTTCC CTCAAGGAAC AATTTTGGTA GAGAAAGGTC CAGCTACTCG 3960

TATTTGGCAA GTTGGCCAGT CTGATTATTA CCAGTTAAAG GTTGGCGGTC CCTATCAGGC 4020

GACTGGTGGT ATGGGTGATA CACTGGCTGG AATGATTGCA GGATTTGCAG GCCAATTTCG 4080

ACAGGCCAGT CTCTACGAAC GTGTGGCAGT AGCAACCCAT CTTCATTCAG CCATAGCCCA 4140

AGAACTATCT CAAGAAAATT ATGTGGTCTT GCCGACGGAA ATTAGTAATT GTCTTCCTAA 4200

AGTAATGAAA AGATATGTCT AAAATAGTTA GACAAAAAAT GTTGATAATT TGTATCATTA 4260

TTCTTAATTC ACAAAAAACG AACGTTTAGT ATTCTTCTTG CTAAGAAACT AAATTTGTTC 4320

GTTTTTTTAC TCTTGTAAAT CTATTTTTGT TAGAGTTGAT TTGGTTTACA TCCGTACTTA 4380

AATTGATTTG TTAGAGCTCT ACTTTTATTA AAAAAATTCA ATTTCAAGGA TAAATAAGCA 4440

GTATTCTAAA GGTACTTTTA GATGAAATAA AAGCCTTTAC ATGGTATAAT AGAGGTAGCT 4500

CTTTAATGGA GGTGTTTGAG TGGAAAATCT GAAGAAAATG GCAGGTATCA CGGCTGCTGA 4560

ATTTATCAAG GATGGGATGG TTGTAGGGCT AGGAACAGGT TCTACTGCCT ATTATTTTGT 4620

CGAAGAAATC GGTCGTCGAA TCAAGGAAGA AGGCTTGCAG ATTACAGCTG TGACGACTTC 4680

TAGTGTGACC AGTAAACAGG CTGAAGGGCT CAATATCCCG CTCAAGTCTA TTGACCAAGT 4740

AGACTTTGTC GATGTGACAG TCGACGGGGC GGATGAAGTG GATAGTCAGT TTAATGGAAT 4800

CAAAGGCGGT GGTGGTGCCC TTCTCATGGA AAAGGTGGTC GCAACACCAT CAAAAGAATA 4860

CATTTGGGTG GTGGATGAAA GCAAGCTGGT CGAAAAACTA GGTGCTTTTA AATTGCCAGT 4920

AGAAGTGGTT CAGTATGGTG CAGAGCAGGT CTTTCGTCAT TTTGAACGAG CTGGCTACAA 4980

ACCAAGTTTC CGTGAAAAAG ACGGCCAACG TTTTGTGACC GATATGCAGA ATTTTATCAT 5040

TGACCTCGCC TTGGATGTCA TTGAAAATCC AATTGCTTTT GGACAAGAAT TGGACCATGT 5100

CGTTGGTGTT GTGGAGCATG GTTTATTCAA CCAAATGGTG GATAAGGTAA TCGTTGCTGG 5160

ACGAGATGGA GTTCAGATTT CAACTTCAAA AAAAGGAAAA TAGAAGGGGG CATAAGATGT 5220

CTAAATTTAA TCGTATTCAT TTGGTGGTAC TGGATTCTGT AGGAATCGGT GCAGCACCAG 5280

ATGCTAATAA CTTTGTCAAT GCAGGGGTTC CAGATGGAGC TTCTGACACA CTGGGACACA 5340

TTTCAAAAAC AGTTGGTTTG AATGTCCCAA ACATGGCTAA AATAGGTCTT GGAAATATTC 5400

CTCGTGAAAC TCCTCTTAAG ACTGTAGCAG CTGAAAGCAA TCCAACTGGA TATGCAACAA 5460

AATTAGAGGA AGTATCTCTT GGTAAGGATA CTATGACTGG ACACTGGGAA ATCATGGGAC 5520

TCAACATTAC TGAGCCTTTC GATACTTTCT GGAACGGATT CCCAGAAGAA ATCCTGACAA 5580

AAATCGAAGA ATTCTCAGGA CGCAAGGTTA TTCGTGAAGC CAACAAACCT TATTCAGGAA 5640

CGGCTGTTAT CTATGATTTT GGACCACGTC AGATGGAAAC TGGAGAGTTG ATTATCTATA 5700

CTTCAGCTGA CCCTGTTTTG CAGATTGCTG CCCACGAAGA CATTATTCCT TTGGATGAAT 5760

TGTACCGTAT CTGTGAATAC GCTCGTTCGA TTACCCTTGA GCGTCCTGCC CTTCTTGGTC 5820

GCATCATTGC TCGCCCTTAT GTAGGTGAAC CAGGTAACTT CACTCGTACG GCAAACCGTC 5880

GTGACTTGGC TGTATCTCCA TTTTTCCCAA CTGTTTTGGA TAAATTGAAT GAGGCTGGTA 5940

TCGATACTTA TGCTGTGGGT AAAATCAACG ATATCTTTAA CGGTGCTGGT ATCAACCATG 6000

ACATGGGTCA CAACAAGTCA AATAGTCATG GAATTGATAC ACTATTGAAG ACTATGGGAC 6060

TTGCTGAGTT TGAAAAAGGA TTCTCATTCA CAAACCTAGT TGACTTTGAT GCCCTTTACG 6120

GCCATCGTCG TAATGCTCAC GGTTACCGTG ATTGCTTGCA TGAGTTTGAT GAACGCTTAC 6180

CTGAAATTAT CGCAGCTATG AGAGAGAATG ACCTTCTCTT GATTACTGCG GACCATGGAA 6240

ATGACCCAAC GTATGCAGGA ACGGATCACA CTCGGGAATA TATTCCATTG TTGGCCTATA 6300

GCCCTGCCTT TAAAGGAAAT GGTCTCATTC CAGTAGGACA TTTTGCAGAT ATTTCAGCGA 6360

CTGTTGCCGA TAACTTTGGT GTGGAAACTG CTATGATTGG GGAAAGTTTC TTAGATAAAT 6420

TGGTATAAGA TGACGCGCTA TGCTTTGCTG GTGAGAGGTA TCAATGTTGG TGGTAAGAAT 6480

AAGGTCGTCA TGGCGGAGCT TCGTCAAGAA TTGACAAACT TGGGACTGGA AAAGGTTGAG 6540

AGCTACATCA ATAGTGGCAA TATTTTCTTT ACTTCGATAG ATTCCAAAGC CCAATTGGTT 6600

GAAAAGCTAG AGACTTTCTT TGCAGTCCAT TATCCATTTA TTCAGAGCTT TTCTTTACTG 6660

AGTCTAGAGG ACTTTGAGGC GGAACTTGAA AATCTACCAG CTTGGTGGAG CAGAGACTTG 6720

GCACGAAAAG ATTTTCTCTT TTACACTGAG GGTTTGGATG TGGACCAAGT CATCGCGACA 6780

GTTGAAAGTT TAGAGCTGAA AGATGAAGTG CTTTATTTTG GAAAACTTGG GATTTTCTGG 6840

GGGAAATTTT CTGAAGAATC CTATTCTAAG ACTGCCTATC ATAAGTACTT GCTGAAGGTG 6900

CCTTTCTACC GCCACATTAC TATTCGTAAT GCTAAAACCT TTGACAAAAT TGGTCAAATG 6960

CTAAAAAAAT AATAAAGGAG ACACACAATG ACATTTTTAA ACAAAATCCA TGAAACTGCT 7020

ACTTTCCTGA AAGAAAAGGG AATTGCAGCC CCTGAGTTCG GTCTAATCCT TGGATCAGGA 7080

CTTGGAGAAT TGGCAGAAGA AATCGAAAAT CCAGTTGTAG TAGACTATGC TGAGATTCCA 7140

AACTGGGGCC GTTCAACAGT AGTCGGTCAT GCTGGTAAAT TGGTATATGG TGAACTGGCA 7200

GGTCGCAAGG TCTTGGCTCT TCAAGGGCGT TTCCATTTCT ATGAAGGGAA TCCTCTGGAA 7260

GTGGTGACTT TCCCAGTTCG TGTGATGAAA GTTCTTGGAT GTGAAGGTGT TATTGTAACC 7320

AATGCAGCTG GCGGTATCGG ATTTGGTCCT GGTACCTTGA TGGCTATCTC AGACCATATC 7380

AACATGACGG GGCAAAATCC ATTGATGGGT GAAAACTTGG ATGACTTTGG CCCACGTTTC 7440

CCAGATATGT CTAGGGCCTA CACACCAGAA TACCGTGCCA CTGCCCATGA AGTGGCTAAA 7500

AAACTTAATA TCAAGCTTGA TGAAGGTGTC TATATCGGAG TTACTGGTCC GACTTATGAA 7560

ACACCAGCAG AAATTCGTTC CTATAAGACA CTGGGAGCAG ATGCAGTTGG TATGTCTACG 7620

GTTCCTGAAG TTATCGTGGC AGCCCACTCT GGCTTGAAAG TTCTGGGAAT TTCATGTATC 7680

ACTAACTTTG CGGCCGGTTT CCAAGAAGAA CTCAATCACG AAGAAGTTGT AGAAGTGACT 7740

GAACGTGTTA AAGGTGATTT CAAAGGCTTG CTTAAAGCGA TTCTTGCTGA ATTGTAAGAA 7800

AAAAGATTTA AAAGGGGGAG TGCCTCTGTT TTTTCAGGAT TGACTGCCTA TCCGGATTAA 7860

AGAAGAAACA GAGGAATACT ATGAGCTTCT TCCTGCTCTT ATAACTGAAA GAAGCGGAAG 7920

AATAGGTATG TCTGATCTGA TAGCCAGCAT TGTGAAAGAC AAGATTCTAG GATACTAGCA 7980

TTAGCTTCCT AGCCAAGCAG ACTAGTATGA TAAGGAGAGA TGAGAATGAA TTGACTTTCT 8040

GAATTTCTCA GTCTTATCAT ATATAGCACA ATGAGATTTC GCTTGAGTCT GCTTGTAAAT 8100

AAACGAAAAG AAAGATAAGA AATAATGAAA ATTGGTCAAC GAATTATGCG CTTTGGCATA 8160

AAAAATTAAG TATCGGAGTT GTATCTGTTG TAGTCGGCTT TGATTTCTAG CTCCAGCTGG 8220

AATTTCAGCC AATGAAGTAA AGCAAGATGT AACATCTGAA GTGGTAATAG GTGTGCTAGA 8280

TTCTAAGGAG GAATTGAAAG AGTCAGAAAA TGATGCTCCA AAACTAGAAA CTCCTCTTAG 8340

AGAGGAGCCA AGACTAGCTC CTCAAACGCT TCCGGAAGCA AGTGAAGTTC TTGAAAACAA 8400

AAGGGAAGAG TCAAAAGTAG AGATAACATA ACCAGCTCAA GCGGATGATA TCCGCAAGGT 8460

TGTTGGGGAA TTAGCCAAGG ATATAAGTAT TACTAAGTTG TATATGACAG GTCATTCTCT 8520

TGGATGTTAC CTAGCTCAGA TTGCAGCGGT TGAAGCTTAC CAAAAATATC CTGATTTTTA 8580

TAACCATGTA TTGAGGAAAG TGACAACTTT CAGTGCTCCT AAAGTGATTA CTTCCAGAAC 8640

TGTTTGGAAT GCTAAGAATG GTTTCTGGGA TGTTGGTTTG GAAAGTCGTA AATTAGCTGT 8700

TAGTGGAAAA ATTAAGCATT ATGTGGTTGA TAATGACAAT GTTGTGACTC CCTTGATTCA 8760

TAATAATCGT GATATTGTTA CATTTACAGG TAATTCACGC TTTAAACACC GTTCTCGTGG 8820

CTATTTTGAA AGTCCAATGA ATGATATTCC TAACTTTAAT ATTGGTAAAC AAGCTACCTT 8880

GGATAAACAT GGTTATCGTG ATCCGAAATT GGATAAAGTG CGATTCTTTA AGAAACAGGC 8940

TCTGCCTCGA TCTTCTAGTC AACCAAGCGC TGAACCAATG GAAAATATTG CCTCAGGAAA 9000

ACAGGTTACT CAAAGTTCGA CAGCTTTCGG AGGAGATGCT AGAAGAGCTG TGGATGGCAA 9060

AGTCGATGGT AACTATGGTC ACAATTCTGT CACTCATACA AACTTCCAAT CTAAGCCTTG 9120

GTGGCAAGTA GATTTGGCTA AAGAAGAAAC CATTCGCCAA ATCAATATTT ACAACCGAAC 9180

AGACACTGCC CAGGATAGAT TGGCAAACTT TGATGTCATT CTTTTAGACA GTTCTGGTAA 9240

AGAAATTGAG TGAAAACGTA TAACATCTCC TAAAGATGTG TCAGCACAAA TTACGATTAA 9300

CCATAAAAAA GCGCGCTATG TTCGGATTGA GCTAGAAGGC TATAATGCCC TCAGTCTTGC 9360

AGAAGTTGAA GTTTTCTGCT TTATAGCTAC GAATGCTGAA ACGGCGACAC AAGTTTCTAA 9420

GCCAGTTCAA CCAATCAGTC AGACTCCTGT GAAGGATAAA ACATTGACAA TTCAACACAG 9480

TGGAGCTTAC ATTGCCCGCT ACTCCATAAC TTGGGAAGAA GTTCCAGTAG ATAAAGATGG 9540

AAACCAAGTT GTTCGTAGTC ATTCTTGGGA AGGAAGCGGT CGCAACCAGA CTGCAGGTTT 9600

TGTCCTCAAC CTCCCAATCA AAGAAAATAT GAGAAATCTG CGAGTTAAGA TTGAGAAAAA 9660

GACGGGCCTA CTATGGAATA GATGGCAAAC AATCTATGAA AACAGACCAA TTTTAGCTCA 9720

ACCCCACCGT AAAATTACCC ATTGGGGTAC GACATTGAAT TCCAAGGTGA GTGACGATGA 9780

TGTCTTGTAA TCTGATGGTA GAATGACAGT TAGTTTGTCT AGTTTATAAG AAAGTACTAC 9840

CTGAGCTTGA ATAGGACTCA GGTAGCTCTC TATGAAAGAA CAAAATTAAT ACTCAATGAA 9900

AATCAAAGAG CAAACTAAGA AACTAGCCGC AGGTTGCTCA AAGCACTGCT TTGAGGTTGT 9960

AGATAAGACT GACGAAGTCA GTCACATATA TAATCCAAGG CGACGTTGAC GTGGTTTGAA 10020

GAGATTTTCG AAGAGTATAA ACAGAAAGGT AGAGCGCGTG TTCTAATTTG AACACGAGTA 10080

GAAAACTTTT CTAAAAACAA AAACGAAAGG ATGGGTAAAC TGTATTCGCT GAACTGAATA 10140

CGGGCGACTC TCCTCTAAAT CAAAATTAAG AAAGGAATTG ACCCCACCCT AAAAGTAGTG 10200

GGAAAAAGAT AGTTGATCTA GCGAGCATCG CTCACTGCGC CCAACTCCTA TTTTCCCTTC 10260

GCTTTTTGAT GGGTTTGGTA TCTTTCTCAA TATAAAATAT AAAATAAAGA AAGGTAGAGC 10320

GTGTGTTTTG ATTTGAACAC GAGCGGAAAA CTCGGAAAAT AGATAATCTG ACTGAAAAAT 10380

CAGGATTTCT CGTCAGGTTC CTAATTTTCA GTCGTTTTCT TCTCGCTCTT TGTATCATAA 10440

ATTATGTCTA TCCATATTGC TGCTCAGCAG GGTGAAATTG CTGATAAAAT TCTTCTTCCT 10500

GGGGATCCTC TTCGTGCTAA GTTTATTGCG GAGAATTTCC TTGATGATGC TGTTTGTTTT 10560

AACGAAGTGC GTAACATGTT TGGTTACACT GGTACTTACA AGGGTCACTG TGTATCTGTC 10620

ATGGGAACTG GGATGGGAAT GCCATCTATT TCGATTTATG CGCGTGAGTT AATCGTAGAC 10680

TACGGTGTGA AGAAATTGAT TCGTGTGGGA ACTGCAGGTT CTTTGAATGA AGAGGTTCAT 10740

GTTCGTGAAT TAGTTTTGGC GCAGGCGGCT GCAACCAACT CAAACATCGT TCGTAATGAC 10800

TGGCCACAGT ACGATTTTCC ACAAATTGCT AGCTTTGATT TGCTTGATAA AGCCTACCAT 10860

ATCGCCAAAA AACTTGGTAT GACTACTCAC GTTGGGAACG TTTTGTCATC TGATGTCTTT 10920

TACTCAAATT ACTTTGAAAA GAATATCGAG CTTGGTAAAT GGGGAGTCAA GGCTGTGGAA 10980

ATGGAAGCAG CAGCTCTTTA CTATCTTGCT GCCCAATACC ATGTTGATGC GCTAGCTATC 11040

ATGACCATCT CTGATAGCTT GGTCAATCCA GACGAAGACA CAACTGCAGA AGAACGTCAA 11100

AATACCTTCA CTGATATGAT GAAGGTTGGT TTGGAAACCT TGATTGCAGA ATAATTATAG 11160

CCAAAAAGGG GCTCTTTGTC AACTGTAGTG GGTTGAAAAA AAGCTAAGCT TGAGAAAGGA 11220

CAAATTTCGT CCTTTCTTTT TTGATATTCA GGGCGATAAA AATCCGTTTT TTGAAGTTTT 11280

CAAAGTTCCG AAAACCAAAG GCATTGCGCT TGATAAGTTT GATGAGATTA TTGGTCGCTT 11340

CCAGTTTGGC ATTAGAATAG TGTAGTTGAA GGGCGTTGAC GATTTTCTCT TTGTTCTTTA 11400

GAAAGGTTTT AAAGACAGTC TGAAAAAGAG GATGAACCTG CTTCAGATTG TCCTCAATGA 11460

GTCCGAAAAA TTTCTCAGGG TCTTTGTTCT GAAAGTGAAA AAGTAAGAGT TGATAGATCT 11520

GATAGTGGTG TTTCAAGTCT TCTGAATAGC TTAAAATCTT GTCAAGAATT TCTTTATTTG 11580

TTAAGTGCAT GCGAAAAGTA GGGCGATAAA AACGTTTATC GCTsArTTTA CGACTATCCT 11640

GTTGGATGAG TTTCCAGTAA CGCTTGATAG CCTTGTATTC ATGAGATTTT CGTTCAAACT 11700

GATTCATAAT TTGAACACGA AAACGACTCA TGGCACGGCT GAGATGTTGG ATAATATGGA 11760

AACGATCTAG AACGATTTTA GCACACGGAA AAAGCTGTTT AGCCAAGTCA TAGTAAGGAC 11820

TAAACATATC CATCGTAATG ATTTTCACTT GACAACGAAC GGCTCTATCG TAGCGAAGAA 11880

AGTGATTTCG GATGACAGCT TGTGTTCTGC CTTCAAGAAC AGTGATAATA TTAAGATTAT 11940

CAAAATCTTG CGCAATGAAA CTCATCTTTC CCTTAGTGAA GGCATACTCA TCCCAAGACA 12000

TAATCTTTGG AAGCCGAGAA AAATCATGCT CAAAGTGAAA GTCATTGAGC TTGCGAATGA 12060

CAGTTGAAGT TGAAATGGCC AGCTGATGGG CAATATCAGT CATAGAAATT TTTTCAATTA 12120

ACTTTTGAGC AATTTTTTGG TTGATGATAC GAGGGATTTG GTGATTTTTC TTTACCAGGG 12180

GAGTCTCAGC AACCATCATT TTTGAAsAGT GATAGCACTT GAAACGGCGT TTTCTAAGGA 12240

GAATTCTAGA AGGCATACCA GTTGTTTCGA GGTAAGGGAT CTTAGACGGT TTTTGAAAGT 12300

CATrTTTCTT CATTAGACTT CCACAATCAG GGCAAGATGG AGCCTCATAA TCCAGCTTAG 12360

CGATAATTTC TTTGTGGGTA TCCATATTGA TGATATCTAG AATCTTGATG TTTGGGTCTT 12420

TAATATCGAG CAGTTTTGTG ATAAAATGTA ATTGTTCCAT ATGATTCTTT CTAATGAGTT 12480

GTTTTGTCGC TTTTCATTAT AGGTCATATG GGACTTTTTT TCTACACAAA AATAGGCTCC 12540

ATAATATCTA TAGTGGATTT ACCCACTACA AATATTATAG AGCCCAAAAA GGAAGCCCTT 12600

TATGAATTGT AGGACTTCCT TTTCTTATCC AGAAATTGAT CTAGCTCTCT CTGATTTCGA 12660

AGAATAGTGA CTTTATGTGA ATATTCTTGG CAAAGTTTTT GGTAATTTTC TTTTTGAGTT 12720

TTGCGGACGC CCATCCCAAA GAATCCATCT GATAAACTCC CACTCAAAGC GTTCAGGGCA 12780

ATCTACCGCC ATACTTTCTC TGACTTTTCC ACGGTATTTA AGATAACGCT TAAAGGCTCT 12840

AAAGAGACAG GTCAATGGCG AAAAATTGAG AAAGATGATT TGGTCAGCTT CTTGCATTCG 12900

TTCTTGGTAG TAGCACCAAG AATAATTACC ATCGATGACC CAAGCTTTAT GCTTGGTGAG 12960

AAAGTTTTTT ATCTCGGTTA ACATCCATTC GCAGTCACTG TCTTGCCAAC CAGGTTGAAA 13020

TTGGAGTGTG TCCATGTGCA GTTTTGGAAT GGAGTAGTAG TTAGATAACT TTTCTGCTAT 13080

AGTTGACTTA CCAGAACCAG AATATCCGAT AATTGCGATT TTCATTTTCT ACCTTTTCCT 13140

ATTTGGAGAC AAAAAAACAG CCTCTATGGA CTGTTTCTTA TTTAACAAGT TTAGCTGAAA 13200

GACGAGCTTT ATCGCGGCTT GCTTTGTTTT TGTGAATCAA ACCTTTAGTT TCTGCTTTAT 13260

CGATAGCTGA GCTAGCAGCA CGGAAAAGTT CTTCAGATGG GTTTGCTTCG AAAGCTTTTA 13320

TAGCAGTACG CATAGCTGAT TTTTGAGCTG AGTTCTTTTC GATTCGTCTA ACGTTCAATT 13380

CAGCGCGTTT GATAGCTGAT TTAATGTTTG CCAATGGTCT TACCTCCATA TTTACTAACT 13440

(2) INFORMATION FOR SEQ ID NO: 129:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8512 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

CCTTTTTTCA AAAACTAGAT ACTAGTCTAT CAAAAGTAGG AAAGGGTTTC AAGAAAATTG 60

ATTGGAAATT TTTTGAAAAT CATAGAACTA TTAGCTAATC CCTAGTATTG AAAAGACTGG 120

ATAGCTTCTT TCAGGTCATC TTGTAAACTA TTTCTCTGGT CAAGTTGGAC ATAGACTTCC 180

ACCAGACAGG ATCTAAAGTT GGAAAATTTG TAAAAATCCT CCCTTTCTTC TATCGGAAAA 240

TCAACAGTTT TTATCCAAGA AGCTACTTGT TCTTGCTCCA ACTTCCCTTG TAAAATAGGT 300

TCATAGATCA CTCTTGCTAA ACGCCAATCC TCATCATCTG TAAAGCGAAT CGACATTCTT 360

TTAAATAGTT GGCCAAGTAT ATCAAATACT TCATGAACTC TGTTTTTAGG AAAGTCTGGA 420

TGACAAACCA CCTCTGTCAG TAAATCGGCT CCATGTGCAA AAGCGTGAAC CCAACCATAC 480

TGACTTGAGA AACCCCTTGT ATCCTTTTCT TTTGAAAGAT AGTGCAAGCC TTGATTTAAA 540

AGGACATTAC GAATTTCTGG AGAAGGATTT CCCAAATGAT CAAACAACCA CTGGATTTCT 600

TCCTGGTTAT AATTTGGTTT TTCTTCTGCT ATTTTTCTTA GTAAATCTTG ATACATGGTC 660

AATACCTCTA CATTTCTAGC AACTGTTCAA AAAGGCAGTC TTAAATGACT CAATATTGAA 720

TTCTCAATTA AATACAATCT GATATAAAAT GACGTAAATA ACTATCAATA CCAGTTCTAC 780

AGTAAGTTCA AATTTAACAT CACGACCTTC AACGACATTT TTGAAAATAG CTACAACTAA 840

GACAAATAGA ATGACGCTTA ACAAGCCCAT AAACATCATT CTAAAAAATT TTTCTATTCC 900

CCTACTCTCC CAACTCAGCA CTATAGGAGA TAATCTGGTC AACTGTGTCA GACAAGAATT 960

GGATGGTATC ACGGAGTGGT TTGTCTGTTG AAATATCAGC ACCGATAATC ATGGCTGACT 1020

CAAGTGGTGT CTTGCTACCA CCTGATTTGA GGAGATTGAG CCAGTCTTCA GCTCCAGTTT 1080

CAGAATGTTT TAGATGAAGG TAACCAGCAG TCGAGATAAC TAGTCCTGCT GAGTAAGTGT 1140

AACTATACAA GCCCATATAG TAGTGAGCTT GGCGCATCCA AGTCAGAGTT GCATCATCGT 1200

CAATTTCAAT AGCATCTCCC CAGAAATCCG TCAAAACTTC CTTCATAATG CTGTTGAGCT 1260

TGCTTGCTCC AAAGGTCTCC CCTTCTTCAA TCAATGTATA CACCTTACGC TGGAAGGCGG 1320

CTTCCAAGAG GTGGGTGATA AAGTTATGGA AGTAGGTGTC TGTCAAGCGA TGAGCCAGAG 1380

CGAAGCGTTT TTGACGTGGG TCATTAGACT GGTTCTCCAA GTAATCACTG AGTAGCAATT 1440

CATTGAAGGT TGACGGTGCT TCAACATAGT AGGTCGACAT ATGGGCATTG AAGTAACTTT 1500

GATGATTGTC TGAAAAGATG AATTGACCAG AATGCCCGAT TTCATGAATC AAGGTATAGA 1560

CATCGCTCAA ACGGCCTGTC CAGCTCATGA GTACATAAGG GTGTACGCGA TATGGGTCCG 1620

CCGCATAACC ACCGGAATCC TTGCCACTGT TAGCAGCAAA GTCCACCCAG CGCTCTTCTT 1680

GGTAACGAGC AACTTCCTGA CAATATTCTT GCCCCAAAGG TTCTACCGAC TTCATGACCA 1740

AATCATAGGC ATCGTCAATA GTCACTTCAG GATTCAGGGC GCTGTCCAAG TCCAATTTCC 1800

AGTCTGCAAA GGTCATCTTT TCAAGACCAT TTACCTTGGC AACATGCTTG AGGTATCTCT 1860

GAGCGACTGG TGCAAAGTCC TTCATGATGA GGTCAATCTG GCGGTCAAAC ATGACACGGT 1920

CCACTTCTTG TTCAGCTAGA AGATAGTCAA AGACAGAGTC GTATCCCTTC ATATCAGCCA 1980

AGAGTTTTTC AGACTTGACC TGAGCCAGAT AGGCTGCTGC AGCCGTATTT TGGTGCTTAC 2040

GAAGTCCCTC TGAGAAGGAA CGGAAGGATT TCTCACGAAC CTCAGCATCC TCATGGTTTT 2100

GGTAGAAATT CTCATAGGTC ACAAAGCTGT TTTTGTAGGT CTTGCCATGG GCTTCAAAGT 2160

CAGCCATTTC AAAATCCCCA GCTCGCATCT TAGTATAAAT GTCCTGCGGA CTGTAGAAAA 2220

CTTCACCGAG ATTTGTCAAG GCCTTCTCCA CATCTGCCCC TAAGTAGTGG GCTTTTTTGA 2280

TTTTAGCCTG ACGAATGGCA GCTGTTAAAT GTGGCAATTT ACCCAAACGG TCCAAGACTT 2340

CCTCATCTGC TGCCACCAAG GCATCGTCAA AGAAGGTCAA GGCTACGCTG GCATCTGTTT 2400

CAAATTCCAT CCCAGCTTGG GCAATATTGG CAAATTCGTC ATTGCTATAG TCCGTCGTCT 2460

GAGGCATAAA ACCATAGTTG CCAATATGGC TCATCTGAAT GTAGATCTGT TCCAATTCCG 2520

CAAAGGCCTT CTCGAAATCC TCAAAAGTGT GAAGATTGCC CTTGTAATCA CGGCTAAACT 2580

GGTTGATGTC TTCGCGAGCT TTCTCGATTG CACGCAAGAA ATCCTCACGG TCTTGGTATA 2640

GGGCTGTTAA GTCCCAGAGT TCCTTCTCTG GAAATTCTGA ACGGTGTTTT TGTTCCATTT 2700

TCTTCCTCTT ATTTCTCTAA TTCTACTAAA ACACTAAGGG CTGATAAAGC GTAAAGCGGT 2760

GCTGTTTCTG CTCGCAAAAT ACGAGGACCT AGGCCTGCCA AAACGGCTCC TTTAGCTTCA 2820

AAACTTTCGA TTTCTGCAGG TGAGAGACCG CCTTCTGGAC CAAAGATAAA GAGCAGTTTG 2880

GCTCCTGTTT CAAGACCAGT GACTGCTTGC AGAAGCGCAG CGGCTTCTCC TTCTTTAGCT 2940

GATTCTTCAT AGGCTACTAT GATAGAGTCA AACTGGTCCA GCTGAGCTAG AAAATCTGCT 3000

TTTTTCTCGA AAAGTTTAAT ACTTGGTACA ATATTACGCT TGCTTTGCTC GGCTGCTCCA 3060

AGGGCAATTT TTTCTAGTTT TTCAACTTTT TTACCCAATT TCTTGCCATC CCACTTGGCA 3120

ACTGACCAGT CTGCAGGAAA GGCCCAGATT TGGCTAGCCC CCAGTTCGGT TACTTTTTGA 3180

GCGATGAACT CCAGCTTGTC TCCCTTGGGA AATCCAGATG CGATGGTCAC TTGGACTGGT 3240

AGTTCCACAT TGTCATTTAA TTCTTGGACC AACTCAAACT GACGATTTTC CATATCCAGC 3300

ACGCGCGCCA AGCGCTTGAT GCCATCATCA AAGACTAAGG TAACCTCATC CTCTTCTTTC 3360

AAGCGCATAA CCTGAAACAT ATGCTTACTG GTTTCCTTGT CCTCGATAGT GACAGGAGAG 3420

ATAGCACTGC CTTTTACAAA ATACTGCTGC ATGCTAGCCT CCAATCACAC CAGAGATATC 3480

CTTGGTTTTC TTAAAGACAC AGGTATTCCA TTCCCCTTGA ACCATGTGAG TTTCGAGGAA 3540

AAATCCAGCT GACTCAGCCG ACTGGCGCAC CATGTCCAAC TTGTCCTTGA TAATGCCACT 3600

CATGATCAGG TAGCCTTCAT CCTTTACCAA GCGATAAGCA TCGTCTATTA GATGAATGAG 3660

GATATCCGCC AAGATATTAG CCACAATCAC ATCTGCCTCA ATTTCCACAC CCTTAAGCAA 3720

ATCTCCAGCC GCTACATGGA TATTTTCCAT GCCAGGGTTG AGCTCAATAT TTTCCTGAGC 3780

CACACGAACC GCCACATCAT CCAGGTCATA GGCGAAAATT TCTTTAGCCC CCAGAAGCGA 3840

GCTGGCAATA GAGAGAACCC CTGAACCAGT CCCCACATCT AGCACCGTTT CGCCACCACG 3900

AAGAACCTGT TCCAAGGCAA AAAGGCTCAT CTTGGTAGTT GGGTGGGTTC CAGTACCAAA 3960

AGCCATGCCA GGATCCAGCT TGATAATCAT TTCCCCCGCA GTCGCCTCAT AGTCTGTCCA 4020

AGAGGGAACG ATGGTCAAAT CATGAGTGAT ACGAGCAGGT TCATAGTATT TCTTCCAGTT 4080

GTCTGCCCAG TCTTCCTCAG CCAAGGCAGT CGTACCTATT TTTAACTCTC CCAAATCCAT 4140

AAAATCTGTC AATTCTGCTA GACGAGCCTG CAAATCCGCC TCAACCACTG TCACATCCAC 4200

CGTGTCAGGG TAGTAGGCTG TCACTACGAT TTCTTCTTGC TGCTCCACCT CTGGGAAAAT 4260

CTCTCCAAAG CGGTCCACAT TTCCCACATA GTCCATACTG TCTTCGATTG CGACTCCTTG 4320

CGCTCCCAGC TCAATCAAGA GATTGGAAAC CAACTCCTCT CCCTCACGCT TCACTGTAAC 4380

TTTTAACTCT TGCCATGTTT CCATTATTAA TACCAAGCCC GTAAAACACA AAACCAAAAT 4440

AGGAAATTCT CTGAAGACGC TTGTGTCTAA GAGAAGTTTA TCTTTTTGGC ACAGTGTTTA 4500

GGGCGGGTTC AGTTTAGAAA TGTAACTGAA CCATCCTTTC TAATCACTTA CTTTTAAATA 4560

ATCTTTTAAT CTCTCTTGCA ACTGAGGCAC AACTTGACTG GAACTAAGAA ATTCCTCAAC 4620

ATTCATCAGC TGATAGCCCT GTCCTTCATC TCCGAAGATG ATATTGTCAA ATTGTTCTTG 4680

TCTTAGCTGA CCAACCATAA AGACCGATTT CTTGCCTTTA AAAATTACGC TAGGATAAAT 4740

CTTGCTCCAA AGCAGACAGT CTTCATCTAA ATGAATTCCC AGTTCCTCAT AAACTTCACG 4800

CCGAGCGCAT TCAAAAGGGC TTTCGTCCCC TTCACGGCCA CCACCTGGCA GTTCCCACAT 4860

ATTGGCCCAG GGAATACTTG CCTTATCATC GCGTAAGATA GTCAAAAGCT TATCCCCACA 4920

AAACAAAGCA ATCTTGCAAC CTGTGAAATC AGAAATTTCT AGTTCCATCT TCAGTTCCTT 4980

CTAACATTTC CTTTTCCAGC TCGGCTAACC AGTTTTCATA ATATCTTTTC TCATCCCTCA 5040

ACATTCGACT ACTATCCATT TTCTGTCTAG CAATCTTGAG AGCCTTACGA GTTCGATCTA 5100

CATCTTTCTT CACCTTTAAT TGATACCAGG CTTGTATCAC TTGAAGATTG GACAGTTTGA 5160

GAGACAGAAA CGATTTGACC TGTCGAATAC TAGCATATTG CTCCGCTTGC TCAAAATCTC 5220

CTTCCAACAA GGCGATATGA AGCAGGGATA GTTGGGCAAC TGTCTGCATC ATCGGAGTAG 5280

TTGTCCTCTC AAGTAATGCT TGAAACTGCT GTTTAGCTAC TTCTTCCTTC CCTTCCAAAA 5340

TGGAAACTTC ACCTTGCATA CCTAATACAC CATCCGCAAA ACTCCCTCGT GCATCCTCAG 5400

GAACTGCTTG AACAAAGTCT TTCAAATCAT ATTCTTGAGG AGCTAGCAAG GTCTGGGCAG 5460

AATGTCTCAA TACCAGGTAG GCGTATTTGG TATTTTCAGG GTGTTGTAGT AATTCCCAAA 5520

TTTTTGCTCC ATCGGTGATG TCGACTGGCA AAATGTTATT TAGGAAGAAA GATAAATTAA 5580

GAAAAATCCA AGTCCCTGCA AAATACCAGC TTCTTGTCAA AAATCCAAAC AATATCGCCA 5640

ATAATATCAA GCCGAGATGA ACCATCAAGC CTCCTGAAAG CATCAGGATG ATTCTTTGAT 5700

CGCTTTCATC CTCTTTTAAA CCAATGTATT GAGCACCAAC ATTTTTCAGA ATGGCTGTTC 5760

TACTAAGATG AAACCTGCCT GACTTTTTGG TCAAAATAAA ATGTCCTAAT CCAAAAGCCA 5820

CCAGCCGATA GCCTGTCAAG TAGCCACAAA AAGCATGACC CAGCTCATGA AGAATAAAGA 5880

TTAAATACAT GCTTAGAAGA GCGAAGGCAT AACCAAAAGT AAAGGCTAAA ACTGCGGAAT 5940

ACCCCAACTC TGCAAATGCG ATTGTTCCAC AAGCAAAAGC TAGCATAATA AAGACAACAG 6000

CTAGCACATA AACCAAATAA GTCCCAATTT TCTTCATAAC ACCTCCAACC AACTCCTAGT 6060

ATCTTGGATA AGGATAAAAT TCTCCCTTTT CCAAGCCAAT TTTTCCTTCT TCAAAGACTT 6120

CTTGGTTCCA TTCCATGACA AATTCCTCTG CTTCTGGGTC TTCCAAAAAG TCCATGAGGA 6180

CATCTAGCCC AACCTCAGCA GTATCTTTAA GGAAAAGCGC AAAATAAGCT AAAAATTCAC 6240

GGGAAAATCC TTTTTTAGGC AGGTAAGGAA TAACAGTCAA ATAGTCTTCC TCATTGACTG 6300

TTGACTTGGC AGGATTGTAG AAAAGGACCG CTTCCTCAAA AAGAATGTCA TCTGATGAAA 6360

CCTCTCCGTC TTCATCCACC ATCTCCACAC CGCAGCATTT TGCGCTTCCA ATAGAAAACT 6420

CACTTCTACC GCATGGTTGC GTTTGTCCCA GCTAATCTCA AAGTCAAAGG GAAAGTTCTT 6480

GTCCAACTCT TCCTCTAAAA TATCTAAAAA TCCGTATGTT GCCATTTTGT CCTCTTTCTA 6540

TGCGACTCTT TAATCGCCCC GATTGCTCGG AAATATGCTA AAATAGATAC TACCATCTTA 6600

CCACAAAATT ATTTTATGTC CTAATTATAC CATATTACCT CATTTAAACC CTTGGTATCA 6660

GTGATTTTCT TAAAAGTCTG ATTTCTTCAT TTCTCATAAA AATCAATATA AAAAGCCCTC 6720

GAAAGGGCTA ATAAATCTAT AAAATCAATA GGCGAGTAAC TAGCACAAGT GGACGTGCTT 6780

TTTTATTGAC TATTACCACG ATACCACGCT TAATCTTAGG CTTGAACTTT CTTATCTGCA 6840

ATAGCGTCTG TCAAAGTCTG AGAAAAGTTA AGCCCCATTT CTCGTCCCAA CTTATCTGCC 6900

CATTTTGGTA TGGTCAAAGT CTTTTTAATG GGTTCCTGAC TTCCTAGGTA TTCTGATACA 6960

TCAACAGATA CCATAGAAAT AAAAGATTTA TCAAGGTCAT AGGTTGACAC GAAATCTTCA 7020

TCATCTTTAA AAGGATCATT ATCAATTAAA GACAAGCTAT TGATATCTGA TGGCTGAGGT 7080

AACTCTCCAT CACTCTCTAT CAAATCTGCA ACAGTTATCC CTAGCCACTC CGACCCCATA 7140

GCCAAAGCCT CAGAAATCCC CTCTCCTTGT GTAGCTGAGT ATTCAAAATC TGGGAAATGG 7200

ACAAAATAAG TCGCTTCTGT TCCGTCTGTG TCGTCATAAT AAAATAAAGC TGGATACGTA 7260

ACTAACATTT CACTACCTCC ATATCAAAAA GCAGGGACTG AATTTTACAA CCCAGCTTGC 7320

TTTCTTATCC CTCTTTCAGT GTACTTATTC AGCTCACCAT GAAGGATTGT GATAGGTCTT 7380

TCCCCTTGCT TTTCCATTTT AATATGGGAG CCTTTACCGC CTCTAGTCTT TATCCAACCA 7440

TGGGCCGTAA GGAGTTTAAC CATCTCTTTT TGTGTCATAG GCATAGCGCT TTTACCTCCT 7500

GACAACACCA TTATAACACG TGTTACACGT ATTGTAAAGG AGTGATACTT ATTATTCTAT 7560

TATACATAAA AGCCCCTAGA TGTGGTTCTA AGGGAAGCCA ATTTATTCAT ACCTATTTTT 7620

CTAATGAGTA GTAAAAACTG CTTCTTTATC GAGCAATTCA TCATCTGTAT AGTCAATTGT 7680

AAAAGTATCT CGATCTAAGA CAGATTGAGG CGGAGTTGAA TGAATCATAG GAACACTGCG 7740

TACTCTATAT TTTTTATCTC CAATTTTTAC AAACTGATAC TCTTCGAAAA TCAAATTCAA 7800

ACCACGTCAA CGTCGCCTTA CCGTACTCAA GTACAGcCTG CGGCTAGTTT CCTAGTTTGC 7860

TCTTTGATTT TCATTGAGTA TGATTAACTC TCAAGTCTTC GAAATCAGGA TTTTCAACAG 7920

TTATTACAAG GAGGCGATTT ACTACTTCAA AAACATCAAT TATTCTATTT TTCATATTTT 7980

TTCAACCCAT TATTAGAATG AACTTCTTGG TAAGCAAAAT CAAGTTTAGA TTTAATGTTT 8040

TCGTACAAAT CTAAAATCTC TTTTGGAGTA TCTTCCCGGA AGAAAAGTTT TCTTTTCCCT 8100

GAAATAACTT GATCACTAAG AATCCAATGA CGAATTTGTT TTGTAAAAAT CAAAATTTCC 8160

TGACTTGGTA GTTCCATCAT TTCCATTGCT TATCACCTCT CTTTTCATTA TAGTTCATAC 8220

AATGACATTC AGCAATATTA TTTCTCAAGT CAGCACTTCC ACTTCTTTAG GCTCAACTAT 8280

CCTATTTTGA GCTTTAAGGA AAATCAAATC TCTCATGCTG ATACCTCTCC TCATTAAATT 8340

AAATAGTAAA AAAGATTCTA TCTCACTCCC TGATTATTAC AAAACCATTG AAATATCACA 8400

ACTAATAGGC TAGAATGGAC ATAGTAAGAT ATAGTAGATG AGTCATTCTA CTCAAATCCA 8460

CGTTAGAAAG GACTGCTATG CCAGACAATC TCGCCGTTCG CATGCGCCCn GG 8512

(2) INFORMATION FOR SEQ ID NO: 130:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2869 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

CTCGTTTCAA GGTTGAGTCT CTTGCAAATC TTGTTCGCGT TCTTCCTTTT GCCAAGGCAT 60

CTCTCCCATG GTTGGTGCcA GCCATTGTTG GAATCTTGCT CTCATTGGTT CTACCAAACA 120

AGCAAGAAAG CGATGTTTTT GAAATGGAAT AATCACTTAA ATCACTTTTG TAGCCAAGTC 180

TACAGGAGTG ATTkTCTTTT TTTATCCGAT GATAAATGTG TTATAATAGG TAGCGAAAGA 240

GGTGAAGAAA TGAATCAAAC AGTAGAATAT ATCAAAGAAC TGACAGCCAT TGCGtCGCCA 300

ACAGGCTTTA CTCGTGAGAT TGCGGACTAT TTAGTCAAGA CTCTAGAAGG TTTTGGTTAC 360

CAGCCGGTTC GCACATCCAA GGGCGGTGTC AATGTAACTA TTAAAGGTCA AAATGATGAG 420

CAACATCGCT ATGTGACTGC CCATGTAGAT ACGCTTGGTG CTATTGTCCG TGCTGTCAAA 480

CCAGACGGCC GTCTCAAAAT GGACCGTATC GGTGGCTTTC CTTGGAACAT GATTGAAGGA 540

GAAAACTGTA CCATTCATGT GGCTAGCACA GGTGAAAAAG TATCAGGAAC CATCCTCATC 600

CACCAAACTT CTTGCCATGT CTATAAGGAT GCAGGAACTG CAGAACGCAC GCAAGACAAT 660

ATGGAAGTGC GTTTGGACGC CAAAGTAACT AGTGAAAAAG AAACTCGTGC TCTTGGCATT 720

GAGGTCGGTG ATTTTATCAG TTTTGACCCA CGAACTGTCG TGACAGAGAC AGGTTTTATC 780

AAGTCTCGCC ATTTGGATGA CAAGGTCAGT GCGGCGATTT TGCTCAATCT CCTTCGCATT 840

TATAAGGAAG AGAAGATTGA ATTGCCCGTA ACAACTCATT TTGCTTTTTC AGTCTTTGAA 900

GAAGTGGGAC ACGGTGCAAA CTCTAACATT CCTGCTCAGG TAGTAGAATA TCTGGCTGTG 960

GATATGGGAG CCATGGGAGA TGACCAGCAA ACAGACGAAT ATACAGTGTC TATCTGTGTC 1020

AAGGATGCTT CTGGACCTTA TCACTATGAC TTCCGTCAAC ATTTGGTGGC TTTGGCGAAA 1080

GAGCAAGATA TTCCATTTAA GCTGGATATC TATCCATTTT ATGGTTCGGA CGCTTCAGCG 1140

GCTATGTCTG CAGGGGCAGA AGTCAAACAC GCCCTTCTCG GTGCTGGTAT AGAGTCTAGC 1200

CATTCCTATG AGCGTACCCA TATTGACTCG GTGATCGCAA CAGAACGAAT GGTCGATGCT 1260

TATCTTAAGA GCACGTTGGT GGACTAATAT GTGCCTTATT TGTCAGAGAA TTGACCTCAT 1320

CAAGAAGGAA GAAAATCCTT ACTTTGTCAA AGAGTTGGAA ACAGGCTATC TTGTGGTTGG 1380

AGACCACCAG TATTTTGAAG GCTATAGTCT CTTTCTAGCC AAGGAGCATG TCAGCGAATT 1440

GCACCATTTG AAAAAGGAGA CAAGACTCCG TTTTCTAGAA GAAATGAGTT TAGTCCAAGA 1500

GGCAGTTGCC AAGGCCTTTG CTGCTGAGAA AATGAATATC GAACTGCTAG GAAATGGCGA 1560

TGCTCATCTT CATTGGCATC TGTTTCCACG ACGGACAGGT GATATGAATG GTCATGGTCT 1620

CAAGGGTCGT GGACCAGTCT GGTGGGTTCC CTTTGAAGAA ATGACAGCAG AAACCTGCCA 1680

AGCAAAACCG GATGAGATTA AAAGATTAGT CAAACGTTTA TCGTCAGAAG TAGATAAACT 1740

ATTAGAAATA AAGGAGTAGA AATGAAGAAA AGATACCTAG TCTTGACAGC TTTGCTAGCC 1800

TTGAGTCTAG CAGCTTGTTC ACAAGAAAAA ACAAAAAATG AAGATGGAGA AACTAAGACA 1860

GAACAGACAG CCAAAGCTGA TGGAACAGTC GGTAGTAAGT CTCAAGGAGC TGCCCAGAAG 1920

AAAGCAGAAG TGGTCAATAA AGGTGATTAC TACAGCATTC AAGGGAAATA CGATGAAATC 1980

ATCGTAGCCA ACAAACACTA TCCATTGTCT AAAGACTATA ATCCAGGGGA AAATCCAACA 2040

GCCAAGGCAG AGTTGGTCAA ACTCATCAAA GCGATGCAAG AGGCAGGTTT CCCTATTAGT 2100

GATCATTACA GTGGTTTTAG AAGTTATGAA ACTCAGACCA AGCTCTATCA AGATTATGTC 2160

AACCAAGATG GAAAGGCAGC AGCTGACCGT TACTCTGCCC GTCCTGGCTA TAGCGAACAC 2220

CAGACAGGCT TGGCCTTTGA TGTGATTGGG ACTGATGGTG ATTTGGTGAC AGAAGAAAAA 2280

GCAGCCCAAT GGCTCTTGGA TCATGCAGCT GATTATGGCT TTGTTGTCCG TTATCTCAAA 2340

GGCAAGGAAA AGGAAACAGG CTATATGGCT GAAGAATGGC ACCTGCGTTA TGTAGGAAAA 2400

GAAGCTAAAG AAATTGCTGC AAGTGGTCTC AGTTTGGAAG AATACTATGG CTTTGAAGGC 2460

GGAGACTACG TCGATTAATA CTCTTCGAAA ATCTCTTCAA ACCACGTCAG CGTCGCCTTA 2520

CCTACTGACT GCGTCGGTTC TATTCACAAC CTCAAAACAG TGTTTTGAGT cGATTCGTCA 2580

GTTTTATCTG CAACCTCAAA GCTGTACTTT GAGCAstGCG GCTAGCTTCC TAGTTTGCTC 2640

TTTGATTTTC ATTGAGTACA AAAAGTAAAC TTTTCTCTTG CAATTCCAGA TAAATAGTGT 2700

ATAATGGATG GGTATGTGAA AAACATACTT GTGGGAGGTA AAAATCTCTA ATTACCGCCA 2760

AAACCACAAA GGAGGATTTA AAAATGGCTA AAAAAGTCGA AAAACTTGTA AAATTGCAAA 2820

TCCCTGCTGG TAAAGCTACA CCAGCTCCAC CGGTTGGACC TGCTCTTGG 2869

(2) INFORMATION FOR SEQ ID NO: 131:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6186 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

CTGAATCCCT TATAGGAGTC CAGTAACTTT TTAGCCTCTA CTTTGCCTTC ATAGGCAGCT 60

TCAACATCAT TAAAAAAAGA ArGCACTGAA GCAAGTTCTT CAGTGCTCCA CGACAAATCT 120

AGTGGGTAAC TATACTGTTT GTTCATTAAC TAATACCAGC TCTCATTCTT GCTTCTTTTA 180

GTTCTTGCTT ACGATAACTA CGAGGGAGAA AAGCACGAAT CTCATCTTCA TTAAAACCGA 240

TTTGCATACG CTTGGCATCA ATAATAATTG GACGACGCAA AAGACTAGGA TACTGCTCAA 300

TCAAATGAAG CAATTCCGAT ACCGAAATAC TCTCTACATC AATATTCAAT TTTTGAAAAA 360

TTTTTGAACG AGTTGAAATG ATGTCATCAG TACCATTTTC GGTCAAGGAA AGGATGTGTT 420

GCAATTCTTT TCTTGTTAAA GGACTGGTCA TAATATTGTG TTCCACAAAG GGAACTTATG 480

TTTTTCTAAC CAGGCCTTAG CCTTACGACA TGATGTACAG CTCGGTGATA GAAATAGTGT 540

AATCATGCTT TTCTCTTCTT ATCTATACTT TGCTACTTCT ATTATACAAA AAAATAAAGC 600

GCTTGACTAG GGATTTTTAG AAAAAAAGCC TATTTTTTCA AGAAAAATAG GCTTTTTGCG 660

AACGATTGAC ACAATTGGAT TTGGTTAATT CACTCTTAAC GATGGTTTTA AACGATATAT 720

ATTTTTATAT ATGTAAATTA AAAACATCTT TCCTTTCACT TCCTACGACT TTTCAGATAC 780

AGATAGCCAA AGAAGTTTTC ATAGAGGGCA AAAAAGAGGA GGAAGGCATG AAGAAAGAAG 840

GTCTCTGGCA AAATCATAAT AACAGGATCC TTGGCTGGAT CAAAAAGCCA GGTATCATCT 900

CCCACAAAGA GAATTTGATG GAAAAGAGTA AAGAATTGGT CAAAACCAAT CAAAACTCCC 960

CCAAGTCCAA TCATCACAGG TAAGACTACT AGAGCCAGGA GACTTTTTCG ATAAAGAGAC 1020

AAAAAGTCCT TTTTCACAAT CCTATTGACA AAGACATAGA AACTTGGCAG TGTCACTAGA 1080

GCTACTAGCT GAACCAAATG AAAGAGATTC TTGACCACTG CGAAATGGTG CAGACCAGCT 1140

GCTGACGAAC GAAAATCAGG CATCTGTAAG ACCTGACTAA AAGGATTGGT CAGATAATTC 1200

ATCAAGATAT GAAAATTGTA TTGAATGGTT TCTGGTTTTA GATAGACTCG ATTCGTTAAG 1260

TTTAGCCACT GAATCTCCAT AGGATAGAAA ATCCAAGCCA GATAAATGGT CAGAAGGATG 1320

GAGAGGGAGA GGAGAAAGAG CATAGAGCCC CAAAAGATCA ATTTAGTTTT CATCAAAATC 1380

CCACTCCGCA AGGCTAGAAA CCACATGTGT CGGTGCGATT GGCAGGCCAG CTACTTCTTC 1440

TGCCTTAGTA AAACCTGTCG TCACCAAGAG CGTTGGAATG CCATTGTCAA TCCCAGCCCG 1500

AATATCAGTC AAATAATTGT CCCCAACCAT GATTAACTCT TCACGTTCCA AACCTAAGTG 1560

CTCAACCGCC TTGTCCATAA TGATGGCATT TGGTTTTCCG ATATAAACCG GCTTCACTCG 1620

TGTCGCTACT TCAAGCAGCG TAATCAGTGA GCCAGCACCT GGCAAAAGAC CGCGTTCCGT 1680

CGGGATGTTG AGGTCAGGAT TGGTTCCGAT AAAATGGGCA CCCTTTTGAA TAGCAAGAGT 1740

TGCTGTGGCA AATTTTTCAT AGTCGACTTG CCAATCCAGA CCAACTACCA CGTAGGCAGG 1800

TTTTTCCTTG TCTTCCACAT AACCAGCCGC CTTGATGGCT TCCTTGAGTC CTGCTTCTCC 1860

GACGACATAG ACGGTCTTTT CAAGCCCCAA ATCATTCATA TAGTCGATGG TTGCCAAAGT 1920

CGCTGTGTAG ACAGTCGATA GGGGCGTATC GATATTAAAA TTCTGAGCCA ACATCTCCTT 1980

AACACTCTCT GGAGTGCGGG TTGTATTGTT GGTTACAAAG AGATAGGGAA TGTCCCGCTT 2040

TTGCAATTCA TGAACAAAAG TCTCTCCAGC AGGGATTCGG TCTTTCCCCT TATAAATGGT 2100

TCCGTCTAAA TCAATTAAAT AGCCTTTATA TTTCATCTAT TTCTCCCTAA GCCTTTTTTA 2160

TTTCTTGCCA AGTAATGATT GCTTGGGCAT TGATAACCCC ATCACTTGTA ATTTCATGCT 2220

TGCTTTCCAG TCCAGTCCGT TCAACAGCCG ATGTAATCAC CCCACCTGGT CGAACTTCCT 2280

TGACATACTT GAGGTTGATT TTCTTGGGAA TATAGTGGGT CAAAAAATCC GCTCCCATGA 2340

CCTCAAAAAT CCAGTCCAAG TATTTACTGT TATTGACATG ACCATTCATA TCCAAGTCGT 2400

AAAAACGAAC ATGGTAATCC TTGCTGATCG GTTCTTCCAA GGACTCATAC TTCGGTCCAC 2460

GGATAAGTTT TTTATCAAAA TCAGACTGGT AAGGAGCCAC AATCTCAGGT TCAACAACAT 2520

GGACTTTTCG ACTGTCGCGG TCCATGAGAA CAAAGGTCGC CATCATGTGG ATGAGCTCCT 2580

GCTCCGCTTC ATTATAAATA GTAAAGCGAC GGTAGCAAAA AAGTCGATTG TAGCTCAAGG 2640

CTTCCGTTTC GATGGTAATT TCTTCCGCAA AACGAGGCAA ACGAACCACC TCAATATCAT 2700

ATTCTACGAT AATCCAGACC AGATTATATT CTTCCAAAAT GGCCTTATCA CTAACTCCCA 2760

GTTCAATCGA CTGCATCCCT GAAACTTGCA GTGACAGCAA AATCACATCT GGAAGTTTGA 2820

TATGACCGTT CATATCAGCC ATATCAAAAG GAATTTTCAT TTTCATTTGA TAAGTTAAGC 2880

CCATGATCCT ACTCCAAAAT AAATCGTTCT GCTACAGTAT CTCCCAAAAA GAGACCTCTC 2940

TTTGTCATGC GAACGTGGTC ACCCTCAATC TGCATGAGGC CTTGTTGAAC CAAATCTCTG 3000

ACAATTTCTC CATAAAGTCC AGCAAAAGAC TGTCCAAATT TTTCCTCAAA TCGCGCCATG 3060

GAAACCCCGG ATTTCTTGCG GAGTCCCAAG AACATTTCTT CTTCCATTTG CTCCTTTTGA 3120

CTCAGGTGAT CTTCTGTAAT ACAAGCATTG CCTTCCTCAA CCGCACTGAG ATAATGACGA 3180

ATGGGACCAT GATTTTTATA GCGTACTCCA TTGACATAAC CAGATGCCCC TGCACCAATA 3240

CCATAGTATT CAGCATTGTC CCAGTACATG AGATTATGAC GACTTTCAAA ACCGGGTTTG 3300

GAGAAATTAG AAATCTCATA ATGCTCAAAA CCCGCTCGCT CCAGCTCTGC AATGATGTAC 3360

TCAAACATCT CCGCTTCTAG TTCCTCCTTA GGCAGAGGCA ATTTCCCACG TCGCATCCGG 3420

TTCATAAAGA CCGTATGGTT TTCTAAAATC AAACTATACA AACTCATGTG GGGAATATCC 3480

AATCCAATGG CTTTAGCCAC ATTTTCCTTT ACTTGCTCCA TGGTCTGACC AGGCAGAGCA 3540

AGGTTTATCC CTAAATCATC CGCTAACTTT CTAGCTGCAG GAGTCGCTCT TAGCTTGTCA 4920

TCAGCCATGA CCTCTCCAAT TCTATTTATG ATACAAAGGG CGTCAAAAGC GACTGAAAAA 4980

TAGGAAATCG ACGATGGCTT CGATGAAGCC AAGGAGATTT ATCTTTTTTC CGATCTTTTA 5040

GCCCGTGCTC TAATCTAAGA TATTAATGAC GAAGAGCTCT GCACCTAAAA GATACAAAGT 5100

TTCTCGTCAG CTTTATTTTA TTTACATAAC TTATCTTATG TAACCCTATT CTTTGTTATA 5160

AGTTTTTCGG ATTGCATCTT TGATACTTTC AACTGTTGGA ATCATTGCAT TTTCTAGGTT 5220

TTGTGCATAA GGCATCGGCA CATCTTCTCC TGCACAACGG CGAATTGGTG CATCTAGATA 5280

GTCAAATGCT TCTGATTCTG AAATAATAGC TGAAATTTCA CCGATATAGC CACTTGTTTT 5340

GTGGGCATCG TTGACCAGAA CAACCTTACC AGTCTTCTTC ACTGAGTTTA TGATGATATC 5400

CTTATCAAGC GGAACAAGGG TACGTGGGTC AACAATTTCA ACTGAAATTC CTTCTTCTGC 5460

TAATTCTTCA GCAGCTTGAA CCACACGGCG AAGCATTTTT CCATAAGTAA CAACTGTTAC 5520

ATCCGTTCCT TGGCGTTTGA TTTCACCAAC CCCAAGTGGA ATTGTGTAGT CTGGATCAAC 5580

TGGCACTTCC CCTTTTTGGT TAAATTCTGA CTTGTACTCA AGTATAATAA CTGGGTTGTT 5640

ATCACGGATA GAAGACTTAA GCAGGCCTTT CATGTCCGCA GGTGTTCCAG GTGCCACAAC 5700

CTTAAGTCCT GGAATGTGAG TAAACCAAGA CTCTAGAGAT TGTGAGTGCT GGGCGGCAGA 5760

GCCAACTCCG TTACCAGCTG CACAACGAAC AGTCATTGGA ACCTGACCTT TACCACCAAA 5820

CATGTAACGT GTTTTAGCAG CTTGGTTGAC GATATTGTCC ATGGCAATAA CAGAGAAGTC 5880

CATGAAGGTC ATATCGACGA TTGGACGAAG TCCTGTCATG GCTGCTCCTG CTGCTGCTCC 5940

AGAGATGGCA GCTTCAGAAA TCGGACAGTC ACGGACACGT TCTGGACCAA ATTCTTCAAG 6000

CATTCCAACA GAAGTACCGA AGTCTCCTCC GAAGACACCG ACGTCTTCTC CCATCAAGAA 6060

CACATTTTCA TCGCGACGCA TTTCCTCAGA CATAGCAAGG ATAATGGTGT CACGGAAGGA 6120

CATTGTTTTT GTTTCCATTT TATCTCTTTC TCCTTAGTCT GCGTAAATAT CTTCAAAGGC 6180

TGATTCAAGC GGTGGGAATG GGCTTTCCTC TGCAAATTTA ACAGAAGCTT CTACTGCTTC 6240

CTTTACTTGC GCTTGGATTT CTTCCAATTC TTCGGCACTT GCAATGTTAT TTTCAATAAG 6300

GTAATTGCGG AGGTTTTCGA TTGGATCTTT TTGTTTCCAC AATTCCACTT CTTCACGCGT 6360

ACGATATTTA CCAGGGTCAG ATGATGAGTG ACCGAGCCAG CGATAAGTTA CACTTTCAAT 6420

CAAGACTGGA CCATTGCCAC TGCGAACATG GTCCACAGCT TTCTGAAATC CTTCATAGAC 6480

ATCGATGACA TTGTTACCGT CTTCGATGAA CATTCCAGGA ATTCCATAAG CGGCGCTACG 6540

TTGATGGATA TGTTCTATAT TGGTCATTTT CTTGATATCC GCAGAGATAC CGTAACCGTT 6600

GTTAATGCAA TAGAAAATGA CTGGCAGGTT CCAGATAGAA GCCATGTTCA CTGCTTCGTG 6660

TCAAGCAATT CAGAACCATT TACCAAGACG TTCTTGTCTT TAGTGATTGT ACCAATTCCT 3180

TTATGAACAG TAACTCCGTA GCTACGAAGA AGTCCTGCAA CACCACCAAC AAGAGTATTA 3240

ACAACTTTAG ATTTAGTTTC TAAAAGTTTT TCCATATCAA CAGTGAAGTT AGGATTTTCA 3300

ATCACGATAC CACGATTTGC AGCATGACCG ATATTTTCAA TAATTTCAGC GTTATGAAGG 3360

TAGGTCTTGG TTGGAATACA TCCACGGTTT AAGCAGGTTC CACCAAGTTC AGATTTCTCA 3420

ACAAGGGCAA CCTTACCGCC GAATTGGGCA GCTTTAATGG CTGCAACATA ACCAGCAGGA 3480

CCTCCACCAA TCACAACGAT ATCAAAAGCA TCATCGCTCT TACCATCATC GTTTGAGGTA 3540

CTTGCTACAG GTACAGGGCT AGCTTCTGGC GATGCTGCTC CAGCTGTTGG GATGTTTTCC 3600

CTTTCTTCAC CAAGGTAACC GATAACTTCC GTTACAGGGA CAGTTTCACC ATCTCCTTTG 3660

AGAATGGCAA TCAAGTACCC ATCTTCTTCG GCTTCCAATT CCATGCTGAC TTTATCAGTC 3720

ATGATTTCCA AAAGGATTTC TCCTTCTTTT ACAAATTCTC CGACTTTTTT ATTCCATTGG 3780

ACGATTTGTC CTTCTGTCAT ATCCACGCCG GCTTTTGGCA TAATTACTTC TAAGGCCATG 3840

TCTTCCTTCC TTTATCTATA TCTTAAAAAT GAATACTCTT GCTCTTAAAT TAACATTGAG 3900

ATTGGCGTTT CAATCAACTC TTTCAAGTCC TTCATAAACT TAGCACCAGC CATACCATCT 3960

ACGACACGGT GGTCAATGGT TAATCCTAAA CTCATGATTG GGCGAATCAC AATTTCACCA 4020

TTGACGACAA CTGGCTTCTC GATTGTCGAA CTGACACCAA GGATAGCTGA GTTGGGTTGG 4080

TTAATAATCG GACCAAAGGA CTGAACACCA AACATTCCCA AATTACTGAT TGTGAATGTT 4140

GAATTTTGTA ACTCACTTGG AGCCAATTTA CCATCCAAGG TACGGCCAAT AACATCCTTA 4200

AAGGCTACAA CCAGTTCTGA AAGACTCATC TTCTCAGCAT TGTAAACAAC AGGTGTCATC 4260

AATCCATTAT CCATCCCAAC TGCCATGGCA AGATTGACAT AGTTGTGAGT GATAATAGTC 4320

TTGCCATCTT CTGTCAATGA AGCGTTGATG TATGGGTGTT TCATAAGAGT CTTAACAACT 4380

GCAAGCGAAA GAAGGTCTGT TACAGTAGTC TTCTTCCCAG TTGCTTCCAT GATTGGCTCA 4440

AGAACCTTCT TACGAAGAGC CAACATTTCA GTCATATCAA CTTCATAGTT GAGGGTGAAG 4500

GTTGGCGCAG TCAAGTAAGA TTCAACCATG CGTTGGGCAA TAACCTTACG CATTGGTGTC 4560

ATTGGAATAC GCTCGATTTT ACCATATGGT GTTACGTTAT CAGGGACTTC TTCCACTTTT 4620

TCAATCTGAG CAGGAGATTT GATGCTATCG TTTTCGATAT TTTCAGGAAG CAGGGCCAAA 4680

ACATCCTTCT TCATGATTTT ACCACGATGA CCGGTTCCTT GGATTTCCTG CCAAGCAATG 4740

TTATGTTCGA GGGCAATTCG TTTTGCAAGT GGCGAAATGC GAACCACGTT TGTGTCTTTA 4800

TAAGTTTCCA CGTCTTCTTT GTGGACACGA CCGTTTGCAC CTGAGCCAGA AACGTCGTAG 4860

CCACGACGGA CGTTAAATTC AGGTGATTTA CCATAGTTCC AGTCCCAAGT TCCAAACTTA 1380

GTATCCTTGA TGCGATTGAT TTCGGCCAAT TCTTCTTCTG AAAAGACGTA TTCAGTCATC 1440

TCTGGGTACT CTTTTTTCAT GTATTCCAAG AGTAAATCAC GGAATTTTTC GACTGTGATT 1500

TTTTTTGGTA ATTCATTGAT AATATTGGTT ACACGGGCAC GGACGGATTT CACACCTTTT 1560

GATTCAAATT TATCTTTTGA AACCTTAAGG GCATTTGCGA GGACTGACAA ATCAACGTCA 1620

AAGAGCAAGC AACCGTGGTG CATGATACGG CCGTTGATAT AGGCTTGGGC ATTGCCACAG 1680

AACTTCTTAC CATCAATCTC AAGGTCATTA CGACCTGTGA ACTCAGCTTT AACCCCAAGT 1740

TGAGCCAGGG TATTGATAAC CGGAGTTGAG AAGCTCTTGA AGTCAAATGC CTTATTTTCA 1800

TCTTCTTTGG AGATGATCGT GTAGTTGAGG TTATTTAAAT CGTGGTAAAC AGCTCCACCA 1860

CCACTAATAC GGCGAACTAC CTCAATACCA TTTTCGCGAA CATAATCACG GTTGATTTCT 1920

TCGATAGTGT TCTGGTGACG ACCAACAATG ATAGATGGCT TGTTAATCCA AAGTAGGAAG 1980

ATTTGATCCT CATCCAAAAG GTGTTTAAAG GCGTATTCTT CCAAGGCAAT ATTAAAAGCA 2040

GTGTCATTTG AATGATTGAT AATGTATTTC ATGATATCCC TTTACTTTAT ATGATAGAAA 2100

CTGGAAATAA CCTTCCAGTC TAATCTATCT TCGTTTTATT TTTTCTTAGG TGAATGGATG 2160

GCCATTCCTA GAACATCTGC AAACGCTTCG TACATCACTT CAGAGTAAGT TGGGTGCCCG 2220

TGGATGGTCT TCAGCATTTC CTCAACAGTG ATTTCCATTT CGATGATGCT TGATGCTTCG 2280

TTTATTAATT CTGCGGCTGC AGGACCAATA ATGTGTACAC CAAGGATTTC TCCGTATTTC 2340

TTATCAGCGA TAACTTTTAC GAAACCTTGA GCTGCGTCAG ATGCAATAGC ACGACCGTTA 2400

GCAGCAAAGT TAAACTTACC GATGGCAACA TCGTATTTCT CACGGGCTTG TTCTTCTGTC 2460

AAACCTACTG CTGCTACTTC AGGGAGAGTG TAGATGGCTG CAGGAGTCAA ATTCAATTTG 2520

GCAACTGCAT GATTTCCTTT AAGGGCATTT TCAGCGGAAA CTTCACCCAT GCGGAAAGCT 2580

GCGTGAGCCA ACATCTTAGT ACCGTTGATG TCACCTGGTG CATAAATGCC TGGAACTGAA 2640

GTTTCCATGT ATTCGTTGAC CTTGATACAA CCACGATCCA ATTCAAACTC AACCTCTCCA 2700

ATACCTTCAA GGTCTGGCAT ACGACCAATT GAAAGAAGAG CTTTGCTTGC GATGATATCG 2760

TCTTTTCCTT CAACCTTGAT ACGAAGTTGA CCATTTTCCT CAATGATTTC TTGCAGTTTA 2820

GTACCAGTCA AGATGGTCAT TCCTTTACGC TCAAGAATCA AGCGAAGGTT CTTAGAAACT 2880

TCCACATCCA TAGCTGGAAC TATACGGTCC ATCATTTCGA TAACAGTCAC TTTTGAACCA 2940

AATGTCATGA AGGCCTGACC GAGTTCGATA CCGACAACTC CACCACCGAT GATAACAAGG 3000

CTTTCTGGCA CTTCGTTCAT TTCAAGAATG TCATCACTAG TCATGACAAG TGGAGATTCC 3060

ATACCAGGGA CGTTGATCTT GTTGACTTTT GAACCACCAG CAAGAATGAT TTTCTTGGTT 3120

CTCCAGG 11887

(2) INFORMATION FOR SEQ ID NO: 147:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11340 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

CCGGTATGTT CTGGAATACT ACCAATCTAA GCTGGCTGTG CCCTACAGTT TTACAACCCT 60

GTACGAATAC CTTAAGGAAT ATGACCGATT TTTCAGCTGG GTTTTGGAGT CTGGTATTTC 120

AAACGCTGAT AAAATATCCG ATATTCCTTT ATCAGTTTTG GAAAATATGT CTAAGAAAGA 180

CATGGAATCC TTTATCCTTT ATCTACGTGA ACGTCCCTTG CTGAATGCTA ATACAACAAA 240

ACAAGGTGTT TCACAGACAA CTATCAATCG AACCTTATCA GCACTTTCTA GTCTTTACAA 300

GTATCTAACC GAGGAGGTTG AAAACGATCA GGGGGAACCT TATTTCTATC GTAATGTAAT 360

GAAAAAAGTT TCCACCAAGA AAAAGAAAGA AACCCTTGCT GCCAGAGCTG AAAATATCAA 420

GCAAAAACTC TTTCTAGGTG ATGAAACAGA AGGTTTTCTA ACTTATATCG ATCAAGAGCA 480

CCCACAACAG CTTTCAAATC GAGCTCTCTC ATCATTCAAC AAAAATAAAG AACGAGATTT 540

AGCCATTATT GCCCTTCTCT TGGCATCTGG TGTTCGCTTA TCTGAAGCTG TTAATCTAGA 600

TCTAAGAGAT CTCAATCTAA AAATGATGGT TATTGATGTT ACTCGAAAAG GTTGCAAACG 660

TGACTCAGTC AATGTCGCTG CTTTTGCTAA ACCTTATTTA GAGAATTATC TGGCCATTCG 720

GAATCAACGC TATAAAACGG AAAAAACAGA TACAGCCCTT TTTTTAACTC TCTACAGAGG 780

TGTTCCTAAT CGTATCGATG CTTCTAGCGT TGAGAAAATG GTTGCTAAAT ACTCAGAGGA 840

TTTTAAAGTG CGTGTAACAC CCCATAAACT GCGCCATACA CTAGCAACTA GGCTCTATGA 900

TGCGACTAAA TCACAAGTTT TAGTCAGTCA CCAACTAGGA CATGCTAGCA CACAAGTCAC 960

TGACCTCTAT ACCCATATTG TTAGTGATGA ACAAAAGAAT GCTCTGGATA GTTTATGATT 1020

TTACGTATTT TAAATTATGT AAATAAATAT CAAAAAAAGA AGTTGGCCAA CTTCTTTTTG 1080

ATTTATCCAA CTACCGCTTC AGCGATTTCT TCACGGCTAA TACCAGCGAA GTAGCGTGTG 1140

ATATCAATGG TTTTTAGCGC CTTAAGAACA TCTTCGCGTT CGTATTTCAC CCCACGAAGG 1200

ACATCTTCTA CTGCAGCAAC GTCTTCAATA CCAAAGAAGT CACCATAAAT CTTGATGTCT 1260

TGGATTTTTG ATTCAGTAAC GTTAGCAAAG ACTTCAACCT TACCACTAGT GAATTTGATT 1320

TTCAACATCT TAAACAGTCG CTGGCTGTGC TCCTGAGGTT AAAAGATACC GTCATACTAC 10140

TGTTTCTGAC GACCAGTATG ATTGCCATCT TGGATGTGTC CCCTCGGCTG ATTGCCCTCC 10200

GCTTCATCCA ACAGACACTA GCACAACTGA GCATTGGGCA ACTCCTCGCC CTGCTCTCCA 10260

TCATCATGTC TTGTGGAGCT ATCCTTGGCA ATATGACCAG CAGTAATCTA TTTAAAAATA 10320

TCCGTTTCAC GCACCTCTTG GTTTTCTGTG AGATTTCCCT ATTGACTCTA ATAACTAGTA 10380

TCCTTTGTCA AGCCTATATC GTAATTTTCA TGACCAGTTT CATCAGTTCT ACGATTATCG 10440

GCATTCTCAG CCCTCGCCTA CAAGCAGCTG TCTTTGCCCA TATCCCCAGT GACAAGATGG 10500

GGACGGTGGG CTCTGCTCTG AGCACAGTGG ACATTCTCGC CCCGTCCCTG CTCTCCCTAT 10560

TAGCCCTATC CATAGCATCG GGCGTTTCGG TGCAGTTAGC ATTGATATTT TTGTATCTTA 10620

TTTTAATTGC TCTTATCTTT TGTCAATGGT TAGTCAAGTT CAACACTCAT AACTAACGAA 10680

AAAGCATGTG TAGATTTCAC ATGCTTTTAA TCTCCCCAAT CGTCAGGTCA AGTACAACAA 10740

AGTCACTTCT TTGATTAAGC GAGTGTTCTA ATATAATTAT AAGCGCCCTG TCATTACCGA 10800

ACCCATTCGC CATTATAGTT GACAGAATAG CCATCTACGG TCGTATTCAC TGCCAAAGCA 10860

CCTGAGCTAT AAGCATAGTA CCAGTTGCCA TTGACCTGGA ACCAACCTGT CTTCATGTCT 10920

CCATTACCTG CATTTAGGTA GTACCAAGTT GAACCATCTT GATACCAACC AGTTGCCATA 10980

GCTCCTGATG AACGGAGATA GTACCATTTG TTCCCAAGGT TTTGCCAACC TGTTTTCATA 11040

TCGCCATTTG GGTGGTCTAA ATAATACCAA GTGGTACCTT CCTGATACCA GCCAGTGGCC 11100

ATTGCTCCTG AGGAACGGAG GTAGTACCAC TTATTACCTA GATATTGCCA ACCTGTTTGC 11160

ATAATACCAG TTGTTGGATC TAGGTAGTAC CAAGTCGAAT CATCGTTTAT CCACCCCGCA 11220

CGTCTTTCAC CACCAAGGTA GTTTTCTCCA TTAATTTCCG TCTTAGCTAG ATAATACCAG 11280

TTAGACTGAT CATAAAGCCA ACCTGTCTCT AAAGAATGAT TTTGATTAAA GTAATAGTTC 11340

GTATAATAAC GCTTCTCTTC TTTATCTTCT GAATCTTCAC GTTTTTCCCC GTACTTTCTT 11400

CCAACACTGT CTTTAGTTTT AATCTCTAAT GTTTTCCAAC CAACAAACTC TTGTAGCACT 11460

CCATTTTTAT CGAAGTAGTA CCACTCTGAC TTTGGAAAAC CTTCTAATCT GATACCATTT 11520

GGGTAAGGAC CAATTGTACT ACCTTTAGAT GGAAACGGGA TATATTGCCA GCCGACAACC 11580

ATCTCTCCAG ATAGAGAATC AAAATAATAG TACTTACCAT CAATCACTCG CCAGTAGGTT 11640

TCTTTGAGGT CCCCCTTTTT GTAGTAGGTT CTTCCGTTTT CTTGGACAAA CTGCCATCCT 11700

TCAGAATCAT CTGCAAATAC TGTACTGGTC CCTAGCAAAC CAAAGAAAAA TACTGTCAGT 11760

CCAACTTGCA TAGTTTTTTT CAAAATTTTC ATCTATATAC CCTCCAATAT TAAATCCACT 11820

CACCAGATGA GGCGAAATTA TAAACTTTAC CATCGATAGT TTGGCTACCT GTAACCATTG 11880

AAATTAAAAA AATAGAGGAA CATAAATATG ATTACAAAAC AGAATGTAAT AGTGTTCTAC 8400

AATTTTTACT AGATAAAACT GTAAATTCTG AAGGAAGGAT CACTTCTTCA ACAGAATTTG 8460

GAAATTTCGT AAGTAATTTA TCATTCCAAC ACGGAATAGC TGGACTACTG TTTCCTCTAA 8520

ATAAATTGTA CCCCCCAGAA CTGGATTCTA AAATACTCTC TATCATCAAG AAGGCAGTGA 8580

CAATTAGAAC GACACACACA TATGAATATC AATACTCACT GCTATTTGGT GATGCAGGCT 8640

ATCTATGGTT ACTCCTACAT TTATTTTCTA TCAGTAAAAA TCAATACTAT CTACAATTAG 8700

CAAACGTCAC CGCTAAAAAA TTAATAGAGA ATTATGATAC TCTAGAGGAA ATAGACTTTG 8760

CATTGGGAAA ATCTGGTGTC CTATTATCAT TAATAAAATA CTATCAATTT ACCAATGACA 8820

ATACTCTTAA AATTTTCATC CACAATAGTA TAGGGGAAAT TTATCATTAT TTCCTACAAA 8880

GAGATACAGC CAAAGAAAGC ATTTTAGACT ATAGCTTTGC TCATGGATAT TGTGGAATTG 8940

CATATGCTTT ATTTGCCTAT TCTAAAGTCT TAGAACCTTC TATGTTTTAT AATGATCTCC 9000

ATACATTCCA TACTGAATTA AAAAAATTAT TAGAAAAAGT TACTTCTAAT ACTGAAAATT 9060

TAGGAAATTT ACAACTTTCT TGGTGCAAAG GAATTTCCGG AATAATCTTA TATCTTTGTA 9120

TGTACGATTG TGACGGAAAC AAAGATATTA TTAGTAAATA TCAAGAATTT GTTTTTAACC 9180

ATCATCTAAA AATGATGACA GGATATTGCC ACGGAATAAC TAGCTTACTA CAAACCACTG 9240

TCTACAATCA AAACAAATTA CTGATGAAAA AAATCCAACA GGTAATTTTA GCATGTTCTG 9300

AACGAGATGA TCACGGTTTA CTGATGTTTC AAGGAGATAG TGGTAAAGCA GATTTGTTTG 9360

ACTTCGGAAT AGGAAGCATG GGGTATATTG GTGTCTATTA AATAATAAAT TCCCATTCGA 9420

TGTGCAGACA TAAGGAGAAA AGTATGAAAT TATTTTGGAC AAACAACATA TATAGACAGT 9480

TGCTGCTAAA CAGCTGTTTT TCATCATTCG GCGACAGTAT TTTCTACCTC GCCATTATCA 9540

ATTATGTGGC TCAGTACAAT TTCGCTCCGC TAGCGATTTT ACTGATTTCC ATTTCAGAGA 9600

TGGTTCCCCT ACTATCGCAA CTCTTTCTCG GGATTCTAGG AGATTTTCAA GAAAATAGAG 9660

TCAAACACGC ACTCTGGATT GCCAAAATCA AAATCCTGCT CTACGCTATT TTGACAGTAT 9720

TTCTCGTCTT GTCGCCCTTT TCATTAGTTT CAGTCATTAT GATTGTCATC ATCAACCTCA 9780

TCTCTGACAC CTTGAGCTAC CTGTCTGCCT ACATGATGAA CGCCCTCTAC ATCAGTGTAA 9840

TTAAGGACGA CCTGCATGAT GCCATGGGGT TCAGGCAGTC TCTGATGAGG GTTGTCCGTA 9900

TTGTCGCCAA TCTGGCTGGC GCATTCCTTA TCAATGTTAT AAGTATTCAA ACTATTTCCC 9960

TTATCAACAC TCTGACTTTT GTCATTGCCT TTTTGGGCCT GTATGTTATT CGACATACCT 10020

TGTATGAGGT TGAAAAAAGA ATTGAAATGT CACATACAGC ACTGAGTTTT AAGAAATATT 10080

CCCAATATTT GATAATAGCA ATATTTGCGT AGGAACGTAC TGTTACAGGC TCTCTATCCA 6600

TGTCTGAACA GCTCCTTTCT CTTCTAATCT TTCTGCTAGT TCTTGTGCGT GTGTCAAATT 6660

GGTTACCAAG GCTATGATAC AACCTCCTAG CCCACCACCG CTCATCTTGG CACCCAGAGC 6720

ACCATGGCTA AGAGTCGTTT CAACCAAAAA GTCTGCCTCA GGGCTACTGA CTCCAATTTC 6780

TTTTAAATGT AAATGCGCTT GACTGAGGAT TTGTCCCAGT CCTTCAGCAT CTTTTTGTGA 6840

AATCGCAACT TCTGCTTGCT GGGTTAATTC TCCCAAGGCA TGCAAAAACG GTAGGGCATC 6900

CTTGCCCTTA TTTTGAACCA CTTGGATGGC TTCACGAGTA TGACCATAAA CACCCGTATC 6960

GGCAATCACC AAATAGGCGG ATAAATCCAT CTCAAGTTCT GTAAATCCTA CGTTCTTGAT 7020

AAAGCGAATA GGTTGGTCAC TAAGACAGGT CTTAGCATCC AAACCACTAG GATTCATATG 7080

GGCAATCATT TCAGCTCGAT TGACCAAGAT TTCTAGTACA TCATGAGGCA GATCAGCCTG 7140

ATAGTAGTCA AATACTGCAC GAATGGCCGC TATGCTGATA GCCGCTGACG AACCCATCCC 7200

CCGTTTCTCA GGGATAGCCG AGTCAATCTC ACAACGAATG CAGGCTTCTG TGATATTCAA 7260

ATACTCCAGT GAGGCATAAA CCGCCATGGA CAAGGTATCC TCCTCATAAA GGCGCCAAGG 7320

ACTCTCTGCA GGAACTACCT TACAGGTCAC CTCCACCTCC AAAAGAGGCA GGGAAATGGC 7380

AGGATAACCG TAAACGACCG CATGTTCCCC TATTAAAATT ATCTTACTAT GTGCCTGACC 7440

GACACCAACT TTTTTTGTCA TTTTTTCCTT TTACTAGACG AAAAAACGTC TTATTTTTCA 7500

TACAAGTATT AATTCTTTCC TATCTATTTT ATTATATTTT CACAAAAAAA GCGATTGTTT 7560

CCATTCACAA TCGCTTCTTT CATTATTGAA CCCATTCGCC ATTATAGTTG ACAGAATAGC 7620

CATCTACGGT CGTATTCACT GCCAAGGCAC CTGAGCGCTA TAAGCGTAGT ACCATCTGCC 7680

ATTGACCTGG AACCAACCTG TCGTCATAGA ACGACGAAAG AAACTCCATA CCATTAAGTA 7740

AAGAGGAAAG TCGTGAGGGA GCATGCGCCA TTGACAACCT GTTTTAGTGA CGTACAAAGT 7800

CTCATTAACA AGTACTCGTT TCGGCCATTT ATAGGTGCGG TGTTTGGAGA AATAGGGTTC 7860

AATCTTCGCC CATTCTTGAT CGTTTAAATC AGTATCATAT GCTTTGCGTA TCATAACTCT 7920

AGCTTAACAT TTTTTTGTGA ATACAGGTTC TAAATAATCG ACCACGAAAA TTTCTTAAGT 7980

GGAAAACGCC TTATGAAGTA TGCTACGGGA AAGTTATGCA CTTAATTTGA CAATTCAAGA 8040

TGTAAAAATA TATACTATAG TAGATTGAAA CTAGAATAGT ACACCTCTAC TTCTAAAATA 8100

TTGTTAGAAA TCGATTTGAC TGTCCTGATC GATTTATCCT GTTATTATCT CATTTTACTA 8160

TAATATTTGA TAAGTTATCC TAAAAGTATT ATTATGTTGT TGTGTTATAG ATTGATTGAA 8220

TCTAACTAAA GGATCCTATT CAATTACTAG AACTATCACA TACTCAAGGT CAGCTCACAG 8280

ATGAGCAACT ATTTTGGTTA CAATGTCTAC TAAATTTAAG TCAAACAAAT AATTTAGTCA 8340

AACGGTCTGC AAATCTTGAC TGGCTTCTTT CAACTGTCTA AGCAAAGGCG TGTAAATATC 4860

TGTACTCAAG CCTTCTAAAA GCTTGCTGGC TACTTCTACT TGATCGATAA TCTTTTCTGA 4920

TTTCCCCTGT TCCAAGGCTT CTACCAGAGA AGTCACCGTT TCTTTTGAGG AAGTTAAAAA 4980

ATTTTGATTG ATATTTTGCT TGATTTGCTG GACCATGTGA CTCGATACAG CCACTTCCTT 5040

GGTCCATCCC ACTAAGAAAT CACATTCTAA AGTTGGTTTC ACTTGTGAAA TTGAAAAGCC 5100

CCAATCACGC TCCAGAACTG TCGCCAAGTT TTCTTCTTCT AACCAAGCAG CCACCTTCTG 5160

GCGATCAAAT GACTGGTAGA GAACCAAATC CTCTGCCACA ATACAGGCAA GGTCGCCCAT 5220

GGAACCATTG TCTCCTCGCT TAAGCAAGAC AGCGCTAGTC AGCTTGAACA AGAGCTCCTG 5280

ATCAACAGAA ACATCATACA GAGCCAGTAA AGCCTTGACA ACCAAGACAA CGACGCTGCC 5340

ACTAGAACCT AGACCAAACT TTTTCCCTTC TCGTTCCATT TTGCCACAGA TTTCTAGAGA 5400

AAAAGGTCTT AAATTCTGAC CACGAACAGC GAGGAAGTCT CCCATCAAAG CAATCGTTTC 5460

TTGAATCAAG CTATAGTCAG GATTAGGCCT TAAGTCCACT GCGAAATCAA ACATATCTGA 5520

ATAGATACGG TAGCTGTCAG AAAAAGCAAT CTCAGCCCTC ATATAGATGG GAATATCCTT 5580

TATCAAAGCT AACTGCCCTG GCTCTAAAAT AGCATATTCA CCTGCCCAAT AGAGTTTTCC 5640

GCAAGTTTTA ACAGCAATCA TCTTGACTCA AATCCTTTGT TTTTGACACA ATCAAGCGAT 5700

AACGATGACC GAAAATTTCT GATAAATGCT CCAAGTCTTT CTCCTGACAG AAGACCTTAA 5760

CATTGGGACC AGCATCCATG GTAAAGTAGC AGGCCTCTCC TTTCTCACGA AGCTGGCGAA 5820

CAAAGGCCAT AGCCTCATAA GAGGCATCCG TCAGATAAGA AAAGGCTGGA CTAGCAGTCT 5880

TTGTCGTAGC ATGCATAGCC AGGGCATTTT TCTCCGTTAA TTCTCCAATC TTGGCAAAAT 5940

CATTTTCCTT GAGATAAATC AGCATATCCT GATAGTCCTT CTCAGACTGA CGAACCCAGT 6000

CGTCGAAAGT CGTCGAGGTT TCCACACAAA GTTTCATCCC GTCACGGCTA GAGATTGGTT 6060

TTTTCTTGTC CTCTAGCACC AACATAATCA TAGCTAGTTT CAAGTCTGTC TCTACAGGGT 6120

AAATTTCTCC ACTATCCTTA TCCCAGGCTC CTAGTGGTCC ATAAAAACTC CGAGAAGAAG 6180

AACCTGAGGC AAATTTGGCT TCCTGTGCCA ACTGACTTCT ATCCAATCCA AGCTTGAAAT 6240

AAGCATTACA AGCCTTGACC AGGGCGGACA AACCACTAGA ACTTGAGGAC AGACCCGCTG 6300

CCGTAGGCAT ATTGTTTTGA GTATCGATAC GGACAAAGCC CTCACCAGCT GGACGATAAC 6360

GGTCAATAAT CTTACTCATC TTGGCATGCT CGACCTCATT TTGTAGCTGA CCATTGATGT 6420

AAAATTCGTC AGCTGTTACA TTGGCTGGTA AAGGCGACAA GGTCGTCTCT GTATACATAT 6480

TTTCCAAAGT TAGAGAAATA CTGCTAGTAG CAGGCACCAT CTCTTTTTCT TTTTTCTTTC 6540

GCTTGGGCTG GTTGAAAAAA ATCAAATCCC CATAGAGACA GTTAACGCTG AGACTGACTT 3060

CCACATCTAC AGGTACGATG ATTTTGGTCG TTCCTACCAT CTTTCTGAGG ATAATGACAT 3120

TGTCATGATT GGTTAAGATG ACCCTCTCCA GATGAATAGT GTCCTTGCCC ATGAAGCGAA 3180

AGAGATTGAT ATCATCGAAT TGGCAAGTCT GGTAGCTTGA AAAATGATGA AGATTTCCAA 3240

ACCAACGATT TTTCTCCTTC TTAACCGTCA CGACCTCTTC AAAAACCAAA TTGGTCTGCT 3300

CTTTTTCCTG GTTCATCATC GGGTAAAGAA GAAAGAGGCT ATAGATAACC GCAACAAAAA 3360

TAGCTAGAAT CACAAAAGGA TTGAGCATAA CGATGAAAAA GAAGAGAATG GTTGCCGCTA 3420

CTAAAAGAAG ATTATTTCCC TCTTTACCAG TGTAGTAGCG AATCAAAAGC AAAAAGAGGA 3480

ATAGTATCAG CAGAAAACGC GAAAAATGCT CTGATACCAT CAAAATCAGA GCTCCTGTCA 3540

GAAGACAGGC TTCGATAAAT AAAAAGATTT TAAATTTTCT CATAGGTTCA TCCTCTCCCT 3600

TCTATTTTAT CACAATTCAA AAAAGTCACC TCAGTCTGAG GATGGAAAAA AGGCGCTGGT 3660

TACGCCTTTT TCATCTGATC CTTTGCTTCT TTTAATTTTC CATAAAGAAG ATAGTCTACT 3720

TTTTGTAGAT CTGCTATGGT GGCACAGTTA AGGGAACACA TAATCAAGCG TAGATCTGCT 3780

TTCCAGCCTT GGACAATGCC AATCACTTCT TCAACTGTGT AGGTTTCAAC CAATTCCAGA 3840

ACGGTTCGTG ACAATCCCAC AGCCTTAGCA CCAAAAACCA AGCACTTAAT CATATCCAGC 3900

GGATTCCGAA CCCCTCCACT AACCAAGAGT TCGACCTTAT CTTTCCATTC TTGGGCATTG 3960

AGAAGGGCCT GCATGGTAGA CTGACCCCAT TGATTGAGGT AATCACGCTG GCCACTACGA 4020

CGGTTTTCGA TATAGGCAAA GCTGGTGCCA CCACGACCCG ATAGGTCCAC TGTACGAACA 4080

CCGAATTCAT AGGCTCTTTC GATTGTCTTG GCATCCATTC CAAAGCCCAC TTCCTTGAGG 4140

ACAATAGGAA CGGGAATTTG CTTGCTATAA TCTGCTAGAT GCGATTGCCA GCTTCTAAAC 4200

TTCCTTTCTC CCTCGGGCAT GAGTAATTCC TGCATGACAT TGACATGCAC TTGCAATAGA 4260

ACAGGATTCA TCTCTTCTAC AGTCTGAAGT CCTAACTCGA CAGGCTTGTC CAATCCAATA 4320

TTGGTTCCAA GGAGGAGATT GGGATGACTA GACTTGACAG AAAAAGAATC ATCCGTTGGA 4380

TTTTTGAGGG CTGCGCTATA AGAACCCGTT ACAAATAAAA TACCACAGGA TTCCGCCACC 4440

TGAGCCAGCT TTTGATTGAT TTCTCTTCCC TTATTACTTC CACCAGTCAT GGCATTGATA 4500

TAAAAAGGAA AGTCCCACTT TCGACCAGCA AACTCTGTCG AAAGATCGAT TTCATCCAGA 4560

TTGTAAAGAG GCAAGGAAGA ATGAATCAGC TCCACCTCAT CAAAGCTATT ATAGGAACTT 4620

TTCTGCTCAA GGGCATAGAG GATATGCTCG TCCTTACGAT TTGTCGTCAT GTCCTATCCT 4680

TTCTTGATAT AAGAGCTCAA TCCCCAGATC GGCCCAACGA TTTTTTAAGG TTTTGGTTGA 4740

TTGCGCATCA AAACTCAGGG CGATGCCACA GTCACCACCA CCAGCACCAC TACTCTTGGC 4800

GCCCCACCAA ATGGTGCTGA AAGGCATAGA CAGCCGCCTG GGTACGATCG CTGACTTCAA 1320

GTTTGGCAAG AATATTGGAC ACGTGGGTCT TGACCGTCTT GAGAGAGATA AAGAGGTCAT 1380

CTGCGATGCG CTGATTTTCG TAGCCCTTGG CGATGAGTTG GAGAACATCT CGCTCACGCG 1440

CAGTCAATTC TTCATGAAGT TCCATATGAT TGCGGTGGTA TTCAACCTTC TTGCTAACCT 1500

CTTGCTCAAT GGCCAGCTCG CCAGCAGCTA CCTTACTGAC GGCATGAAGC AATTCATCTG 1560

CACTAGAAGT CTTGAGCATA TAGCCTTTGG CACCAGCATC TAAGACTGGC ATGATTTTTT 1620

CATTGTCCAA ATAAGAGGTC ACAATCAAAA TCTTGGCTTC AGGCCATTCT TTAAGGATTG 1680

CTAAGGTCGC GTCAATCCCA TTCATCTCAG GCATGACAAT ATCCATGACA ATGACATCTG 1740

GACGCAGTTC CAAGGCCAAG TCAATCCCTT GAGACCCGTT GGACGCCTCA CCCACAACTT 1800

CTACATCGTC TTGGAGGTCA AAGTAGCTTT TCAAGCCCAA TCGGACCATT TCATGGTCAT 1860

CTACTAGTAA AATTTTCATC TTTACTCCTT TATCATTCCT TATCTAACAG GGGAATACGG 1920

ATATCAACCG CCAGCCCTTG CTTGGGAGCT GTCAAGAGTT GAACTGTTCC AGCCATATCT 1980

TCAACCCGCT CCTTGATATT TCGCAGTCCA TAACTCAAGT CGTCTAAGCT CCCTAACTGG 2040

AAACCAATCC CATTGTCCAC CACCTTCAGT TGCAATTCAA CATCTGTCTG ATAGAGGTAG 2100

ACATCTAGGC AAGATGCCTG GGCATGGCGG AGGGTATTGC TAATCAACTC TTGCAGGATA 2160

CGGAAGATAT GCTCCTCGAT TTTCTTAGGC AATTTCGTCA TATTCTGCTT GAGACTAACC 2220

CTAAGATCAC TCTTGTCCTC AAGCTCTTTT AAAAGAATTT GAATCCCTTC TATCAAGCTC 2280

TTCTGCTCCA GTTCAACTGG TCGCAAATGC AAGAGCAAAA CCCGCAAATC CTTCTGGGCT 2340

GTTTCTAAAA TAGCTGTGAC ACTCTGCAAC TGGGTCTGCA TCTTTTCTCT ATCCAATTTC 2400

AAAGCCTGCT GACTGATACC CGATAAAATC ATGTGGGCCG CAAACAACTC CTGACTGACT 2460

GTATCGTGCA AATCCCGAGC AATTCGCTTC CGTTCCTTCT CGATGATTTC CTCTTCCTGA 2520

GCAAGGCTCT GATTTTCAGC TTTTTGAAGA GCCTCTGTCA AAAGGTTAAG TTTACCTGAT 2580

AAGGACTTGA AACTGGCATC CAAATCTGGA TCTGCAACCT GAACCACTTC TTGCCCTGCT 2640

AATAAACGCT TGAGATTAGC CTGCATTTTT CTTAGAGAAA GCTCTTCGAT CCCTCGCCAA 2700

AACAGGGCTA AGAGACAGGT CATGGACATG CTGAAAACCA ACAATAAAAA GACAAATTTT 2760

TCTGTTTTTT CGACATCGTG CAAAAAGATA GACCAGTCAA AATCAAGTAT TTCCAGCAAG 2820

CTGTGGGAGA AAAAAAAGAC AAATAGGAAG GAGGTGAGAG CAATAATGAC ATAGGCTTGT 2880

TTTTTCATCC TCTAACCACC TCCACATCAC CAATCATAGT GGTCAAGAAA ATCTTGACAC 2940

TCTTGTTACT CTTGAGATAG TCTTTTGTTT CTTGATGATA GTGTTCATTG CGGAGGGCTC 3000

CTAACTCACA TTAATTGCGT TGCGCTCACT GCCCGCTTTC CAGTCGGGAA ACCTGTCGTG 10620

CCAGCTGCAT TAATGAATCG GCCAACGCGC GGGGAGAGGC GGTTTGCGTA TTGGGCGCTC 10680

TTCCGCTTCC TCGCTCACTG ACTCGCTGCG C 10711

(2) INFORMATION FOR SEQ ID NO: 146:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11887 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

TACATTCATT CCATCGGCTA CTCCATAATA CTTAGATAAA ACCATAGCTG AAGTCGAATA 60

CGGATACTGT AAAGTATTAT CAATTTTAAT CAAATCATCA TTACCGATAA TACTTCTGAT 120

TGCTTTTGGT AGTATGAACC ATACGTTGGT GAAATCTCAG ATAATGAAGA ATCATTAGAC 180

TCTGGACCTT TTTCTAGTGT CTCACTTACC TCATATTCTT CACCCTTACT AGAAATAACA 240

CTCAAAGCAG ATACTGTCGA TAACTGGCTA GCCAATAAAG TACTCGCAAT AATTGAAATA 300

CCCAATTTTT TATAAACAGT TTTCTTCATT ATTGTATCCT CCTAATGTAA TTATAGCGTA 360

CTATTCTAAA TTTCTTAATC TACTATAGAA TCAAGAAATC TACCACCTTC TTTAAATACC 420

CTCCATTATC ACATAAACAG GTAAACTTTT CAATTAATGA CTGCGCTTTT CAATCACGCT 480

AGAGGTACTT GCTTGCTTCT TTGATACTAA GTTCAGCCAT TCTTTCCTTG TTTTTCTCAA 540

TAAAGCATGT TACCCAAGTG GGATTCGTTT TGGAGTAGTC TCGCAGAGTC CAGCCAATGG 600

CTTTATTGAT AAAAAATTCT GTTTGGTTCA AGTTATGAAG GAGAATCTTT TCCATTAATT 660

GAGTATTGGT CTTCTCTTTT CTTAACAACT GGTGGTCAAT AGCGACACGT CTCAGCCAGA 720

TATTATCTGA TAGGCTCCAT TTTATACTCA ATGAAAATCA AAGAGCAAAC TAGGAAGCTA 780

GCCGCAGTTG CTCAAAACAC TGTTTTGAGG TTGCAGATAG AGCTGACGTG GTTTGAAGAG 840

ATTTTCGAAG AGTATTAAGA TTATTTCTTC TAGTTCAGGG TGTTCATACA CCAAACTCCC 900

TACTACTCGA TCTAGGATAT CTACCGTGTC CCACAAGGAT TTTGTCACGA CTAACTGCTC 960

TAGCTTAGGC AAATCGGTTT CCTTTAGATA AGACTGCATT GCTTTCAAAT AGTTAGCAGC 1020

CACATATTGG TATTTTCTAG GATCCTTTTC CCAGCAAGTG TCTGCAAAAT CCCAATCGAT 1080

AATCTTTGTT TTTTTCGCTT CTGGAAAATA TTTTATAGAG TTTATTTCTT TCAGGCACCG 1140

CAATACCTAG AAAAGAAAAT TGATGGCGCA TATAGGCTTC CATGGACCTT GCTTTTTTAG 1200

AGTCTTTTGC TGCTTCTAGC TCCTCAAGTA AATCTGCTAA ACTCATCTAA AACTCCTCTT 1260

CCCAGTCGCG GTTTGTTTCT ACTGCTTGGT TAATTCCGAA GGAAACATTA CTTGACTGAT 8880

GGCGTGCTCC TAGCTGGGCA ATGACTTTAC CGTTAGAAAC ATCAACAATG GTAGAAGCGA 8940

CTTGCAATTC ATCGTCTGGA TAGGCAACGT ATTCGTCTGT ATTGTAAATA TCCCACAGAT 9000

GTTTTTGAGC TTCTTGGTCT ACATTTGTGT AGACATCCAT CCCAGTTGTG AGTAGGTTAT 9060

AGCCTGTTTC TTCTTCAACT TGATTGATGA CTTCCTTGAG GTAATTATCC ATGTAAGCAG 9120

GGTAATTACT TGCTGATTTG AGACTTTGTA GTCCATCAGT AATTGGTGTA TTGACTGCTT 9180

TCTCATACTG TTCAGCAGAG ATGTAGCCTT GATTTTTCAT TTCAGATAAG ACCAAGTTTC 9240

GGCGGTCTTG GGCTGCTTCT GGATGTGAAT AGGGGTCATA TTGGTTTGGT GCCTGAGGCA 9300

TTCCAGCCAG CAAGGCTAAC TGAGGTAAAC TTAAATTATT GAGGTCTTTA CCATAGTAGT 9360

TTTGAGCTGC TGTCTGCATT CCATAGTTCC CATTAGACAT GTAGACCTTA TTTATATAGT 9420

AGGTCAAGAT TTCTTGCTTG GTTGCTTTTT GTTCTAACTG AATCGCTAAC CAAGCTTCCT 9480

GAGCCTTACG AGAAATAGTC TGGTCGGAAG TCGAAGTTGA AAAGTAAGTC AACTTAATCA 9540

ACTGTTGGGT GAGAGTTGAT CCACCTTGGA GGGAATTGCT TTGCAGATTG CGCAAGAAAG 9600

CTCCCAGGAT ACGGATGGTA TCAATCCCCC TGTGGTCGAA GAAGCGATGG TCTTCGATAG 9660

AAACGATTGC CTTAACCAAA TCTGTGGGAA TATCATTAGC TTGGGCATTG ACGCGGCGTT 9720

CAGAACCCAA GTCAGCAATG AGTTGATTTT TATTGTCGTA GATTTTACTA GAAGTTGTTG 9780

CAACTAGTTT ACTCTCGGAT AGGCTAGGAG CCTTGCTAAC GTAGTAGAAA AAAACTCCTC 9840

CGCCTAAGAC AATGGCTGCG ATAACCAAGC TTAAGAAGCT AATGCTCAGA TACTTGATTA 9900

GGCGCAGAAT CGTTGGTTTG TTCATCTTGT TTTACCACCT AATAAATGTT CTTTGATAAC 9960

ATTGAGATAA GGAATTTGAG GGAAGGCACC AGCCTTGATT TCATATCCAT ATTCTCGAAT 10020

ATATTCAAGT GGCATTGATT TTTGTCCCTT ATCTTGATGA TAGAAGCGAA TCAAATCGAA 10080

TGCCGGCAAT AAGTAGGTTT CTTGCTGAGA AGAAAAGTGA AGAAGGACAA AGCAGATTCC 10140

TTGTTGGGCA AGGACTTGTT CCATATGCTG AATCTGATGT GGATGAAAAT TTTTCATCGG 10200

AATCGCACGT TTTTGTTTTG TTTCCTTGAC TTCAAAGTCG ATGTAATATC CATTATAAAC 10260

GCCAGAATAG TCCGTCGTTG AAGCTTGTCG AAAATAGGCT TCAACAATCT TGGCACGACT 10320

TCGTTGTGGA TAGTCCACTT GTACGATTTG AATAGGAGTT GGTTTCTTAT GTATAACAGC 10380

CAAGCCCTGA GACAAATAGT AGTCGTTGGT AGCATTGATC ATCTTTTCAA AGGGTACCGA 10440

GCTCGAATTC GTAATCATGT CATAGCTGTT TCCTGTGTGA AATTGTTATC CGCTCACAAT 10500

TCCACACAAC ATACGAGCCG GAAGCATAAA GTGTAAAGCC TGGGGTGCCT AATGAGTGAG 10560

TTTACTGCAA CTGATTGGCC AAATAACAAC AAACCAAAAG ATATTGACTT TGGTAAGACA 7080

ATCAAGGCTA AGAAAATTGT CCTTACTGGT ACCAAGACAT ACGGAGATGG TGGAGATAAA 7140

TACCAATCTG CAGCGGAACT TATCTTTACT CGTCCACAGG TAGCAGAAAC ACCTCTTGAC 7200

TTGTCAGGCT ATGAAGCAGC TTTGGTTAAG GCTCAGAAAT TAACAGACAA AGACAATCAA 7260

GAGGAAGTAG CTAGCGTTCA GGCAAGCATG AAATATGCGA CGGATAACCA TCTCTTGACG 7320

GAAAGAATGG TGGAATACTT TGCAGATTAT CTCAACCAAT TAAAAGATTC TGCTACGAAA 7380

CCAGATGCTC CAACTGTAGA GAAACCTGAG TTTAAACTTA GATCTTTAGC TTCCGAGCAA 7440

GGTAAGACGC CAGATTATAA GCAAGAAATA GCTAGACCAG AAACACCTGA ACAAATCTTG 7500

CCAGCAACAG GTGAGAGTCA ATCTGACACA GCCCTCATCC TAGCAAGTGT TAGTCTAGCC 7560

CTATCTGCTC TCTTTGTAGT AAAAACGAAG AAAGACTAGT ATTTAGTAAA ACCTCTTAAC 7620

AAGATTACGG AAGCAGTCTC TATCTTTTCC AATGAGGTTT ATAGTACAGA AAAAGCCTGA 7680

GAAGATGTCT TCTCAGGCTT TTGTTAAGCA CATAAATACA ATAGTGCTAT GACAAAATCA 7740

CCCAGAAAAA TCTGGGTGAT AAATGTTATG GTTGTGCTGG TTGAGGATTC TGATTTTGTT 7800

GATCAGGGGT TGTATTTGAT TGTTGCGTAT TATTGTTAGG ATTGGTAGTC GTACTATTAT 7860

TTGTGCTTGG AGTGGTTGAG CTAGACTGTG AAGTTGAACT ATCTGATGAT GAGCTTGAAC 7920

TTTCAGTTGA TGGGGGTTGT TGTGGAGCAG GTGAGTTCCA CGTAGAACGA GCACCATTTT 7980

TAAATACGAA TTCTCCATTT CTGTAGAGCC CCTCTGGTAT ATTCCAATCT TCTGGATTGC 8040

TTCCTTCAGA CAGGTAGGTC ATCATAGAGC GGTAAACTTT GGCAGCGACC GTAAGGCCAT 8100

TGCCTACAAG TGGTGTCAGA CGGTTAGAAT AGCCTGTCCA TACAGCCATT GAATATTTAC 8160

GCGTATAGCC AGCAAATAGT TCATCAGGTG CTACAAATTG AGAGGTCTTG ATGTGGTTTT 8220

CAATTTCCTC GTCTGTATAG TTAGAGGTTC CTGTTTTACC AGCCTGAGGG AGCCAAGCAA 8280

GATAGGCATT TCGTCCAGTT CCATAAGTCA AGACTGTTTT CATCATGTCG GTCATCATAT 8340

AGGCTGTCGT TTCCTTCATG GCACGAGTTC CGACATTAGA GAACTCTTTT TCACTCCCAT 8400

CACTAAAGAC GACTTTATGG ATATACATTG GTTTATAGTA AGTTCCACCA TTTGCAAAGG 8460

CAGCGTAAGC AGCAGCCATC TTTTCACTAC TTGCTCCATA TTTTTTGTCT GATTCGGTTG 8520

TGTTACTTGA AATGGCATTT GAGTAGTGAA TACTTGGGTA GTCGATTCCT AGACCATTTA 8580

GGAAAGTCTT GGCGCGGTTG AGTCCGACCT TGTTTAGAGT TTCCACGGCT GGGACGTTTC 8640

GCGATTGTTG CAGGGCGTAT TGCAAGGTGA TGTTGCCAAA GTAGCCCCTA TCCCAGTTAT 8700

AAACAGGAGT ATTTGTCCCA GGGTAGTTAT AGGGCTCATC GTGAACGATA GTAGCAGTTG 8760

AATCGTAGAC ACCGTACTCC AAGGCAGGAG CATAGTCTGT GATCGGTTTC ATAGTTGATC 8820

ACCCTTCCAA GCGATTGGGC AAAGAGCAAG GTTTACCTTT ACAAGCTAAC TGACCAAGGT 5340

AAGACAGAAG AGCAAGAACT AACTGTAAAA GATGGTAAAA TTACCCTAGA TCTTCTAGCA 5400

AATCAACCAT ACGTTCTCTA TCGTTCGAAA CAAACTAATC CTGAAATGTC ATGGAGTGAA 5460

GGCATGCACA TCTATGACCA AGGATTTAAT AGCGGTACCT TGAAACATTG GACCATTTCA 5520

GGCGATGCTT CTAAGGCAGA AATTGTCAAG TCTCAAGGGG CAAACGATAT GCTTCGTATT 5580

CAAGGAAACA AAGAAAAAGT TAGTCTCACT CAGAAATTAA CTGGCTTGAA ACCAAATACC 5640

AAGTATGCCG TTTATGTTGG TGTAGATAAC CGTAGTAATG CCAAGGCAAG TATCACTGTG 5700

AATACTGGTG AAAAAGAAGT GACTACTTAT ACCAATAAGT CTCTCGCGCT CAACTATGTT 5760

AAGGCCTACG CCCACAATAC ACGTCGTGAC AATGCTACAG TTGACGATAC AAGTTACTTC 5820

CAAAACATGT ACGCCTTCTT TACAACTGGA GCGGACGTCT CAAATGTTAC TCTGACATTG 5880

AGTCGTGAAG CTGGTGATCA AGCAACTTAC TTTGATGAAA TTCGTACCTT TGAAAACAAT 5940

TCAAGCATGT ACGGAGACAA GCATGATACA GGTAAAGGCA CCTTCAAGCA AGACTTTGAA 6000

AATGTTGCTC AGGGTATCTT CCCATTTGTA GTGGGTGGTG TCGAAGGTGT TGAAGATAAC 6060

CGCACTCACT TGTCTGAAAA ACACAATCCA TATACACAAC GTGGTTGGAA TGGTAAGAAA 6120

GTCGATGATG TTATCGAAGG AAATTGGTCA CTCAAGACAA ATGGACTAGT GAGCCGTCGT 6180

AACTTGGTTT ACCAAACCAT CCCACAAAAC TTCCGTTTTG AAGCAGGTAA GACCTACCGT 6240

GTAACCTTTG AATACGAAGC AGGATCAGAC AATACCTATG CTTTTGTAGT CGGTAAGGGA 6300

GAATTCCAGT CAGGTCGTCG TGGTACTCAA GCAAGCAACT TGGAAATGCA TGAATTGCCA 6360

AATACTTGGA CAGATTCTAA GAAAGCCAAG AAGGCAACCT TCCTTGTGAC AGGTGCAGAA 6420

ACAGGCGATA CTTGGGTAGG TATCTACTCA ACTGGAAATG CAAGTAATAC TCGTGGTGAT 6480

TCTGGTGGAA ATGCCAACTT CCGTGGTTAT AACGACTTCA TGATGGATAA TCTTCAAATC 6540

GAAGAAATTA CCCTAACAGG TAAGATGTTG ACAGAAAATG CTCTGAAGAA CTACTTGCCA 6600

ACGGTTGCCA TGACTAACTA CACCAAAGAG TCTATGGATG CTTTGAAAGA GGCGGTCTTT 6660

AACCTCAGTC AGGCCGATGA TGATATCAGT GTGGAAGAAG CGCGTGCAGA GATTGCCAAG 6720

ATTGAAGCTT TGAAGAATGC TTTGGTTCAG AAGAAGACGG CTTTGGTAGC AGATGACTTT 6780

GCAAGTCTTA CAGCTCCTGC TCAGGCTCAA GAAGGTCTTG CAAATGCCTT TGATGGCAAT 6840

GTGTCTAGTC TATGGCATAC ATCTTGGAAT GGTGGAGATG TAGGCAAGCC TGCAACTATG 6900

GTCTTGAAAG AACCAACTGA AATCACAGGA CTTCGCTATG TTCCGCGTGG ATCAGGTTCA 6960

AATGGTAACT TGCGAGATGT GAAACTTGTT GTGACAGATG AGTCTGGCAA GGAGCATACC 7020

AAGCTTCGCG ATGATGCTCA CTTAATCAAT GCGGAAATGA CAGTACGCTT GCAAGTTGTA 3540

GACAATCAAT TGCACTTTGA TGTGACTAAG ATTGTCAACC ACAATCAAGT CACTCCAGGT 3600

CAAAAGATTG ATGACGAAAG CAAACTACTT TCTTCTATTA GTTTCCTCGG CAATGCTTTA 3660

GTCTCTGTTT CTAGTAATCA AACTGGTGCT AAGTTTGATG GGGCAACCAT GTCAAACAAT 3720

ACGCATGTCA GCGGAGATGA TCATATCGAT GTAACCAATC CAATGAAGGA TTTGGCTAAG 3780

GGTTACATGT ATGGATTTGT TTCTACAGAT AAGCTTGCTG CTGGTGTTTG GAGTAACTCT 3840

CAAAACAGCT ATGGTGGTGG TTCGAATGAC TGGACTCGTT TGACAGCTTA TAAAGAAACA 3900

GTCGGAAATG CCAACTATGT AGGAATCCAC AGCTCTGAAT GGCAATGGGA AAAAGCTTAT 3960

AAGGGCATTG TTTTCCCAGA ATACACGAAG GAACTTCCAA GTGCTAAGGT TGTTATCACT 4020

GAAGATGCCA ATGCAGACAA GAACGTTGAT TGGCAAGATG GTGCCATTGC TTATCGTAGC 4080

ATTATGAACA ATCCTCAAGG TTGGGAAAAA GTTAAGGATA TCACAGCTTA CCGTATCGCG 4140

ATGAACTTTG GTTCTCAAGC ACAAAACCCA TTCCTTATGA CCTTGGATGG TATCAAGAAA 4200

ATCAATCTCC ATACAGATGG TCTTGGGCAA GGTGTTCTCC TTAAAGGATA TGGTAGCGAA 4260

GGCCATGACT CTGGTCACTT GAACTATGCT GATATTGGTA AGCGTATCGG TGGTGTCGAA 4320

GACTTCAAGA CCCTAATTGA GAAGGCTAAG AAATATGGAG CTCATCTAGG TATCCACGTT 4380

AACGCTTCAG AAACTTATCC TGAGTCTAAA TACTTCAATG AAAAAATTCT CCGTAAGAAT 4440

CCAGATGGAA GCTATAGCTA TGGTTGGAAC TGGCTAGATC AAGGTATCAA CATTGATGCT 4500

GCCTATGACC TAGCTCATGG TCGTTTGGCA CGTTGGGAAG ATTTGAAGAA AAAACTTGGT 4560

GACGGTCTCG ACTTTATCTA TGTGGACGTT TGGGGTAATG GTCAATCAGG TGATAACGGT 4620

GCCTGGGCTA CCCACGTTCT TGCTAAAGAA ATTAACAAAC AAGGCTGGCG CTTTGCGATC 4680

GAGTGGGGCC ATGGTGGTGA GTACGACTCT ACCTTCCATC ACTGGGCAGC TGACTTGACC 4740

TACGGTGGCT ACACCAATAA AGGTATCAAC AGTGCCATCA CCCGCTTTAT CCGTAACCAC 4800

CAAAAAGATG CTTGGGTAGG GGACTACAGA AGTTATGGTG GTGCAGCCAA CTATCCACTG 4860

CTAGGTGGCT ACAGCATGAA AGACTTTGAA GGCTGGCAGG GAAGAAGTGA CTACAATGGC 4920

TATGTAACCA ACTTATTTGC CCATGACGTC ATGACTAAGT ACTTCCAACA CTTCACTGTA 4980

AGTAAATGGG AAAATGGTAC ACCGGTGACT ATGACCGATA ACGGTAGCAC CTATAAATGG 5040

ACTCCAGAAA TGCGAGTGGA ATTGGTAGAT GCTGACAATA ATAAAGTAGT TGTAACTCGT 5100

AAGTCAAATG ATGTCAATAG TCCACAATAT CGCGAACGTA CAGTAACGCT CAACGGACGT 5160

GTCATCCAAG ATGGTTCAGC TTACTTGACT CCTTGGAACT GGGATGCAAA TGGTAAGAAA 5220

CTTTCTACTG ATAAGGAAAA GATGTACTAC TTCAATACGC AGGCCGGTGC AACAACTTGG 5280

CAGATAGTGC TTTGATTATT CCAACTACAT CTCGTACAGG GCGTCCAATC TTGTCTAAGA 1800

TGGTACCATT TACAATACCA TTTGCATTGT CAGGAAATAA AGGTACAAGT GAACCAGTCT 1860

TGTATAAATA CTTGGAACTT CAAGACAAGG CAGTCACTGT AGATGAATAC CAAAAAGCTC 1920

AGGAAAAATG GATGAAAGAA AAAGAAGAGT CTAATAAAAA GGCTCAAGAA GATCTCGCAA 1980

AACATGTGAA ATAACTGTTG CAAAATATAA GAAAGGATTT AGTATTTCCC TTGAATGCTG 2040

AATCCTTTTT TACATTTGTA AAGAAAGATT CTAAAATGTA CGGACCCCCA AAAGTTGGAG 2100

CCTCTTTTTG TCAGAATAGA GAAAATTTTT GTTAATTTTA CTTGTTTCCT ATTGCTTTCT 2160

CAGCTATTAT TTGTTATATT AAAAGTATAA TTATTTTTTA TTTATCAGAG TTAAGCATTG 2220

CACTTTCAGA GGAAGGAGTA TTTTTTAAAA AGAAAATGTA AACGTTTGCT CAAAAATGAA 2280

AGGATTTAGA AGTTTATGAA TAAAGGATTA TTTGAAAAAC GTTGTAAATA TAGTATTCGG 2340

AAATTTTCAT TAGGTGTTGC TTCTGTTATG ATTGGAGCTG CATTCTTTGG GACAAGTCCG 2400

GTTCTTGCAG ATAGCGTGCA GTCTGGTTCC ACGGCGAACT TACCAGCTGA TTTAGCTACT 2460

GCTCTTGCAA CAGCAAAAGA GAATGATGGG CGTGATTTTG AAGCGCCTAA GGTGGGAGAA 2520

GACCAAGGTT CTCCAGAAGT TACAGATGGA CCTAAGACAG AAGAAGAACT ATTAGCACTT 2580

GAAAAAGAAA AACCGGCTGA AGAAAAACCA AAAGAGGATA AACCTGCAGC TGCTAAACCT 2640

GAAACACCTA AGACGGTAAC CCCTGAATGG CAAACGGTAG CGAATAAAGA GCAACAGGGA 2700

ACAGTCACTA TCCGAGAAGA AAAAGGTGTC CGCTACAACC AACTATCCTC AACTGCTCAA 2760

AATGATAACG CAGGCAAACC AGCCCTGTTT GAAAAGAAGG GCTTGACCGT TGATGCCAAT 2820

GGAAATGCAA CTGTTGATTT AACCTTCAAA GATGATTCTG AAAAGGGCAA ATCACGCTTT 2880

GGTGTCTTTT TGAAATTTAA AGATACCAAG AATAATGTTT TTGTCGGTTA TGACAAGGAT 2940

GGCTGGTTCT GGGAGTATAA ATCTCCAACA ACTAGCACTT GGTATAGAGG TAGTCGTGTT 3000

GCTGCTCCTG AAACAGGATC AACAAACCGT CTCTCTATCA CTCTCAAGTC AGACGGTCAG 3060

CTAAATGCCA GCAATAATGA TGTCAATCTC TTTGACACAG TGACTCTACC AGCTGCGGTC 3120

AATGACCATC TTAAAAATGA GAAGAAGATT CTTCTCAAGG CGGGCTCTTA TGACGATGAG 3180

CGAACAGTTG TTAGCGTTAA AACGGATAAC CAAGAGGGGG TAAAAACAGA GGATACCCCT 3240

GCTGAAAAAG AAACAGGTCC TGAAGTTGAT GATAGCAAGG TGACTTATGA CACGATTCAG 3300

TCTAAGGTCC TCAAAGCAGT GATTGACCAA GCCTTCCCTC GTGTCAAGGA ATACAGCTTG 3360

AACGGGCATA CTTTGCCAGG ACAGGTGCAA CAGTTCAACC AAGTCTTTAT CAATAACCAC 3420

CGAATCACCC CTGAAGTCAC TTATAAGAAA ATCAATGAGA CAACAGCAGA GTACTTGATG 3480

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

CCGGAGAAAA TGATGAAAAG TTCAAAACTA TTTGCCCTTG CGGGCGTGAC ATTATTGGCG 60

GCGACTACTT TAGCTGCATG CTCTGGATCA GGTTCAAGCA CTAAAGGTGA GAAGACATTC 120

TCATACATTT ATGAGACAGA CCCTGATAAC CTCAACTATT TGACAACTGC TAAGGCTGCG 180

ACACAAATAT TACCAGTAAC GTGGTTGATG GTTTGCTAGA AAATGATCGC TACGGGAACT 240

TTGTGCCGTC TATGGCTGAG GATTGGTCTG TATCCAAGGA TGGATTGACT TACACTTATA 300

CTATCCGTAA GGATGCAAAA TGGTATACTT CTGAAGGTGA AGAATACGCG GCAGTCAAAG 360

CTCAAGACTT TGTAACAGGA TTAAAATATG CTGCTGATAA AAAATCAGAT GCTCTTTACC 420

TTGTTCAAGA ATCAATCAAA GGGTTGGATG CCTATGTAAA AGGGGAAATC AAAGATTTCT 480

CACAAGTAGG AATTAAGGCT CTGGATGAAC AGACAGTTCA GTACACTTTG AACAAACCAG 540

AAAGCTTCTG GAATTCTAAG ACAACCATGG GTGTGCTTGC GCCAGTTAAT GAAGAGTTTT 600

TGAATTCAAA AGGAGATGAT TTTGCCAAAG CTACGGATCC AAGTAGTCTC TTGTATAACG 660

GTCCTTATTT GTTGAAATCC ATTGTGACCA AATCCTCTGT TGAATTTGCG AAAAATCCGA 720

ACTACTGGGA TAAGGACAAT GTGCATGTTG ACAAAGTTAA ATTGTCATTC TGGGATGGTC 780

AAGATACCAG CAAACCTGCA GAAAACTTTA AAGATGGTAG CCTTACAGCA GCTCGTCTCT 840

ATCCAACAAG TGCAAGTTTC GCAGAACTTG AGAAGAGTAT GAAGGACAAT ATTGTCTATA 900

CTCAACAAGA CTCTATTACG TATCTAGTTG GTACAAATAT TGACCGTCAG TCCTATAAAT 960

ACACATCTAA GACCAGCGAC GAACAAAAGG CATCGACTAA AAAGGCTCTC TTAAACAAGG 1020

ATTTCCGTCA GGCTATTGCC TTTGGATTTG ACCGTACAGC CTATGCCTCT CAGTTGAATG 1080

GACAAACTGG AGCAAGTAAA ATCTTGCGTA ATCTCTTTGT GCCACCAACA TTTGTTCAAG 1140

CAGATGGTAA AAACTTTGGC GATATGGTCA AAGAGAAATT GGTCACTTAT GGGGATGAAT 1200

GGAAGGATGT TAATCTTGCA GATTCTCAGG ATGGTCTTTA CAATCCAGAA AAAGCCAAGG 1260

CTGAATTTGC TAAAGCTAAA TCAGCCTTAC AAGCAGAAGG AGTCCAATTC CCAATTCATT 1320

TGGATATGCC AGTTGACCAA ACAGCAACTA CAAAAGTTCA GCGCGTCCAA TCTATGAAAC 1380

AATCCTTGGA AGCAACTTTA GGAGCTGATA ATGTCATTAT TGATATTCAA CAACTACAAA 1440

AAGACGAAGT AAACAATATT ACATATTTTG CTGAAAATGC TGCTGGCGAA GACTGGGATT 1500

TATCAGATAA TGTCGGTTGG GGTCCAGACT TTGCCGATCC ATCAACCTAC CTTGATATTA 1560

TCAAACCTTC TGTAGGAGAA AGTACTAAAA CATATTTAGG GTTTGACTCA GGGGAAGATA 1620

ATGTAGCTGC TAAAAAAGTA GGTCTATATG ACTACGAAAA ATTGGTTACT GAGGCTGGTG 1680

ATGAGACTAC AGATGTTGCT AAACGCTATG ATAAATACGC TGCAGCCCAA GCTTGGTTGA 1740

TGTTGAAGAA ACTTTAATCC AACCAACCTT TGTCTATGGA CATCCAGTAG CTGTATCTCC 1860

ACTCGCTAAG AAAAATCCTG AAGACCAACG CTTTACTGAC CGTTTCGAGC TCTTTATCAT 1920

GACTAAGGAG TACGGTAATG CCTTTACTGA GTTGAACGAC CCAATCGACC AACTTAGCCG 1980

TTTTGAAGCC CAAGCTAAAG CCAAAGAACT TGGTGATGAT GAAGCGACAG GAATCGACTA 2040

TGACTACATT GAAGCTCTTG AATACGGTAT GCCACCAACA GGTGGTTTGG GAATCGGTAT 2100

CGACCGTCTC TGCATGCTCC TCACTGATAC AACAACTATC CGTGATGTAT TGCTCTTCCC 2160

AACAATGAAA TAAATTCTTA TCCTCTGGGT CTTATCAGAG GATTTTTTGA TTCAAAAAGA 2220

GACTGAATTT AAGGAGAAAA TGAAGTGTAG TATATTGAAA TTGAAATAGT ACACTTTGAT 2280

TTCTAAGACA TTGTTAGAAA TTGGTTTAAA TTCCCTAAGC AATTTGTGCA TGTTTTATTT 2340

CATTTTACGA TAGTACGCTG AAACTTTTCA AAAAGTACTA GAAATTGACT TGGATTCCCC 2400

AATTGATTTG TTCAGATTCA CTATAAATAA AAAATTAATA AGTGGGATAG GAAGTTAGCG 2460

TCAACTAGGA TAGTATCTTG CTTAAACAGT ATATATGGGA TTGATATAAG TCCATAGGTC 2520

CTATTAGAGG ATGTTCTGGT GTCTTATTCA CTTGTTTTTT ATAGTATTAG TAGATAGAAT 2580

CAGCAAATAA AAACCCAAAT CATTCATACC TCTCTCAACT AGATGTAACT TACAAAACCC 2640

CTGACCTCAT GAGCCACTTT CTTCCTCCTC ATGAGGTCAG TTTTACTTTC TGCTGTTCCA 2700

GTATCGTTTT TCCTCGCTAG ATTTCCTCAA AAGGGCAGAC TCCTCCCTTG GTGCGTCACA 2760

CGATTTTTTC ATCTCGACTG TTCTTTAATG CATCATTAAC GACGCTTTTC TTCTAGGTGG 2820

TTCATAAGGA ACAGGAAGAT TCAGGTTGAC TTTTCTAATC CTAGAATAAA GTGCTGAAAA 2880

CAATTCGGAA TAGGCATAGA GACTAGACAA TTTGAGGAGC TGCTTGCGTC CTGTTCGAAC 2940

ACATTTTCCC ACCACGTGAA GAAAAAGATG GCGGAAGCGT TTGATTGTTA AAGTTTGGAA 3000

GTCACCTCCA GCTAGATGTT TGAGAAAAAG ATAGAGATTG TAGGCGATAC AGCTCATCAT 3060

CATACGAACT TCGTTTTTGA TTAAGGTTGA ACTATCCGTT TTATCGCCAA AAAATCCCTC 3120

CTTCATCTCC TTGATGAAAT TCTCGGCTTG ACCACGTCCA CGATAAAGCT GAAACTGGTC 3180

TTGGCTTGTT CCACTCGTCA TATTTGTAAC GAGAGAAATA ACATCGTAGA AC 3232

(2) INFORMATION FOR SEQ ID NO: 145:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10711 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

CAGGGGCGTA TTACGTGACA ATTCAATGTA GGCTGTCGCT ACTTGCGCCA AAACAAGGAT 60

TCGATAATGT CGGATGATAC TAACGATTAA ACCGAGCAGA AAGGATCCCA AAATTCCCCA 120

AACTGCAATA TGCAAGGTCA GAAAGAATGC CTTTTGATAT AGTGGTAGAT ATTGTTCAAC 180

AATGGATCAA TCCAAAAATA GAACCTCCCA TCTAGAAATA ATACAGTTAT TGTAGCACTT 240

AAAATCTTCT TTGGATAATA TCTATTTTTT ATTGCCGTTA TAAGGATTTT TATCATAGAC 300

ATAAAATTTC TGAAATTTCC AAACAAAATA TTTTAAAAGT TTTGAAAAAG AGTTAAGATA 360

TTTTTGTAAT ACACAAAGTA AACGCTTACT TATTAAGGAG GACATTTTAT GTCATACAAA 420

ACAAGCAATG CAGAAGGTCA TGTAGATTTC ATCAATACCT ATGATTTGGA GCCAATGGCG 480

CAACAAGTTA TTCCTAAAGC AGCATTTGGC TATATCGCTA GTGGGGCGGG AGATACTTTC 540

ACTTCTTTCC AGTGATTTTA GCGTCAGGTT CTTTTTAGTT TTTAAAGATT ATCCGTGAAT 600

TTCTTGCTTA TTTATGATAA AATGGGAGTG TCGCAAAAAA TGACTCATCG TATTCAATTT 660

TGAGTAAAAC TAGGAGGATC CCATGTCTAC AGAACATATG GAAGAACTAA ATGACCAGCA 720

GATCGTTCGC CGTGAAAAAA TGGCTGCGCT CCGCGAACAA GGAATCGATC CTTTCGGAAA 780

ACGTTTTGAA CGTACTGCAA ATTCACAAGA ATTAAAAGAT AAATATGCCA ACCTCGATAA 840

AGAACAATTA CACGATAAAA ACGAAACAGC TACTATCGCA GGACGCTTGA TAACCAAACG 900

TGGTAAAGGA AAAGTTGGTT TTGCCCACCT TCAAGACCGC GAAGGCCAGA TTCAGATCTA 960

CGTTCGTAAG GATGCTGTCG GTGAAGAAAA CTACGAAATC TTCAAAAAAG CAGACCTTGG 1020

TGACTTCCTT GGTGTCGAAG GTGAAGTGAT GCGTACGGAT ATGGGAGAAC TCTCTATCAA 1080

GGCAACCCAC ATCACACACT TGTCTAAGGC TCTTCGTCCT CTTCCTGAGA AATTCCATGG 1140

TTTGACAGAC GTTGAAACAA TTTACCGTAA ACGTTACCTT GACTTGATTT CTAATCGTGA 1200

AAGCTTTGAA CGCTTTGTCA CTCGTTCAAA AATCATCTCT GAAATCCGTC GTTACCTTGA 1260

CCAAAAAGGA TTCCTTGAAG TGGAAACACC TGTTCTTCAT AATGAAGCCG GTGGTGCTGC 1320

TGCCCGTCCA TTTATCACCC ACCACAATGC CCAAAACATT GACATGGTGC TTCGTATCGC 1380

GACTGAGCTT CACTTAAAAC GCCTTATCGT GGGTGGTATG GAACGTGTCT ATGAAATTGG 1440

CCGTATCTTC CGTAACGAAG GAATGGACGC TACTCATAAC CCTGAGTTCA CTTCTATCGA 1500

AGTTTACCAA GCTTATGCAG ACTTCCAAGA CATCATGGAC TTGACTGAAG GCATTATCCA 1560

ACACGCTGCT AAATCAGTCA AAGGTGATGG CCCAGTCAAC TACCAAGGTA CTGAAATCAA 1620

GATTAACGAA CCATTTAAGC GTGTTCATAT GGTGGATGCT ATCAGAGAAA TTACTGGTGT 1680

CGATTTCTGG CAAGACATGA CTTTGGAAGA AGCTAAAGCT ATCGCTGCTG AGAAGAAAGT 1740

TCCAGTTGAG AAACACTACA CTGAGGTTGG TCACATCATC AATGCCTTCT TTGAAGAGTT 1800

CTGAGCAACT CTTTGCAGGA AACATCAAAA ACCTTTCTGT CAAAGAGCTC AAACAAGGAC 3660

TTCGTGGTGT GCCCAACTAC CAAGTACAGG CAGACGAAAA CAACAATATC GTGGAACTGC 3720

TCGTCTCATC TGGTATAGTT AACTCAAAAC GCCAAGCCCG TGAAGACGTC CAAAACGGAG 3780

CCATCTACGT AAACGGCGAC CGCATCCAAG AGCTTGACTA TGTCTTGAGT GACGCTGATA 3840

AGTTAGAGAA TGAACTGACT GTTATCCGTC GTGGGAAGAA AAAATACTTT GTATTGACTT 3900

ACTAAACTAT TCAACATTTA TCTATAAACA AAGGAGTTAA CCTCGAGAAA GGTAACTCCT 3960

TTTTGCTGTT AATAACTCTC ATCTATCTAT TTTTAATAGA CAGGCTACGC AGGACAATGC 4020

GCAAGGTTGT TAGATTATGT AAGATAGAGA GATTTGAAGG ACTGAACCAA TTAAATAAGC 4080

CAAAGCCAAT CAAACTACTA TTTACGACAA CGGTATCCTG AATATTTTTC TTGATGAGTG 4140

TTTGCAAAGA TGATGATAAC GAATCCAACT CTTGGAAGAA ATCCAAACGA TTATCTAACA 4200

ATAAGATATC ACTCATCTGC TTAGAAATAT CTGCACTCTC ATTCATCACC ACACCGATAT 4260

CTGATAGAGT TAAAGCCGCT GAGTCATTCA ATCCATCTCC AACCATCAAA ATAGTGTGAC 4320

CTGCTTTCTG CAGTTTCTCT ACTAACTCAA ATTTCCCATC AGGTTTCAAG TCTGTATAGA 4380

CCTGATCAAA GGGCAAATCT TTGACTAATT CCTCTGTCCT AATCAAGGTG TCTCCTGTTG 4440

CCAGAATCAA TTTTTTCCCC TGTGCCTTAA GTTTATCCAA GGCTGTTTTT GCTTCTTTTC 4500

TCAAAGGAGT ATGAATGCAG AACATTCCAA TCAATTCATT TTGATAAGCC AAGAATAAGA 4560

GATTGTAGTG ACTCTTGTAC TCTTCAATTA AAGCATTTTG TTCTGAACTG ATATGAATCT 4620

GCTCATCCTG CATCAAGACA TAATTCCCAA TAAGAACTGG TTGGCCATCT ATATGAGATT 4680

TGATCCCCTT GCTTGCGATA TATTGGAGTT TCCCATGCAT TTCCTCATGT TCAATTCCCT 4740

CTATCTCAGC TTGCTTGACG ATGGCATTAG CAATAGGATG ATAAATGTGT TCCTCAAGAC 4800

AGGCACTGAT TCTGAGAATA TCTTCCTCAC TATAGTCTCC AAAAGGTAAC ACCTTTTCAA 4860

CTATAGGATA ACTAGTTGTG ATTGTTCCTG TCTTATCAAA CAAGAAAGTA TCAACTTCCA 4920

GATATTTCTC CCTGTTGTGG CCTCTGGCTG TCATCTCTGT GCTGG 4965

(2) INFORMATION FOR SEQ ID NO: 144:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3232 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

TTCTGTCCCT TATTATTTCG GCCAAAGGGA GCCACATTGA GATAGGTCGT TAAAATCTCA 1860

TCTTTATTCA TGGCGCGTTC CAAGGCAAGA GCATCCACAA TCTCTGCCGC CTTACGAGCC 1920

AAGGTCGGCG CATCCCCAAC CACCTGCTGT TTAATTAGTT GCTGGGTCAA GGTTGAACCC 1980

CCACTAGAGG AACCCAAACC TACAAATTTC CCCAAGGTCG CACGAATCAC CGCCTTGGGT 2040

ACTACACCCT TATGTTCTTT AAAGTGTTCA TCTTCTGTCG CAATGATAGC CTTCTTCAGA 2100

TTTTCCGAAA TTTGCTCAGA TGAGATAGAA GTGCGCAACA AATCACTCTC TATGGAAGCA 2160

ATCACCGTCC CGTCCGAATA GGTAATCTCT GAAATAGAAG AGATGTCCTT GACCTGATTC 2220

ACCAATTCTT CTGTCTGAGG CACCCGAACC TTGTCAAATA AGGCCACTCC GTATCCCAAA 2280

GCAATCCCAG CTCCCAACAT TCCTCCTAGA AAACCGAGTA CAAAGAGTAA GTTAAATAAG 2340

GCTTTTATAC TCAGTAAAAT AGCTGGGAAA ATGACTGACT TATCTAAGGT TTTAGATTTT 2400

TTGGTACTTG AACCTTTCTT GCCAGGTCTA GCTGATTTTT TATTTTTTTG TTTTTGCTGG 2460

AAAAATTCCA GCATTTTTCG TTTTAATTCA TTTAATTGAT TTTGCATGGA TTTCCTCACT 2520

TTATCTATTA TACCACAAAA GGGAAATTTT CAATAAAATA GCCACTTTCT TCCCTATTCT 2580

GCTAGGCTAT TGCCCAAGTT TGTGATACAA TAGGTAGAAA CAATAATTTT AAAAAGGAGA 2640

AAAAACACAT GCACATTTTT GATGAGCTAA AAGAGCGTGG TTTGATATTT CAAACGACTG 2700

ATGAAGAAGC TTTGCGTAAA GCCCTAGAAG AAGGTCAAGT TTCTTATTAT ACTGGCTACG 2760

ATCCAACTGC TGACAGCCTT CACCTAGGCC ACCTTGTCGC AATCTTGACA AGTCGTCGCT 2820

TGCAACTAGC AGGTCACAAA CCTTATGCGC TCGTTGGCGG TGCTACAGGT CTCATCGGAG 2880

ATCCGTCCTT CAAAGATGCT GAACGTAGTC TCCAAACAAA AGACACAGTA GATGGCTGGG 2940

TCAAGTCTAT CCAAGGACAA CTTTCTCGTT TTCTTGACTT TGAAAATGGC GAAAACAAGG 3000

CTGTCATGGT CAACAACTAC GACTGGTTTG GCAGCATCAG CTTCATTGAC TTCCTCCGTG 3060

ATATTGGAAA ATACTTCACG GTCAACTACA TGATGAGTAA GGAATCTGTT AAAAAACGGA 3120

TCGAAACAGG AATTTCTTAC ACTGAGTTCG CTTACCAAAT CATGCAAGGG TATGACTTCT 3180

TCGTCCTTAA CCAAGACCAT AATGTCACTC TTCAAATCGG TGGTTCTGAC CAGTGGGGAA 3240

ATATGACAGC TGGTACCGAA TTGCTTCGTC GTAAGGCGGA CAAGACTGGT CACGTTATCA 3300

CTGTTCCACT AATCACAGAT GCAACTGGTA AGAAATTTGG TAAATCAGAA GGAAATGCCG 3360

TCTGGCTCAA TCCCGAAAAG ACTTCTCCAT ACGAAATGTA CCAATTCTGG ATGAACGTGA 3420

TGGACGCTGA CGCTGTTCGC TTCTTGAAAA TCTTTACTTT CTTGTCACTT GATGAGATTG 3480

AAGATATTCG TAAACAATTT GAAGCAGCGC CACACGAACG CTTGGCTCAA AAAGTCTTGG 3540

CTCGTGAAGT TGTTACACTT GTTCACGGAG AAGAAGCCTA CAAAGAAGCA CTTAACATCA 3600

GAAGAAGGTC GTGTAGTACT TGAGTTACTG CTATCGCTAG AACTACTACT TGAACTGCTG 120

GAGCTGGATG GAGTTGGTAG ACTCCCCACA ATACTAGACC AAGCATTCTG ATAATCCGCA 180

TCACTTCCGC CAATAGCAAA GCGATAACTT GTCGCTGGCG CTCCTGACTT ATTAGCCCAA 240

TAGCTGGTAA CAGTCGAACC TGTGACCTCT ACTTCTTTTC CTTCAACAGA AACCTTCTCT 300

GGTTTTTGAC CTGTTGATTT CAAGACTTCC GATTTCACTA CACTAGGATC TAAAGCAAAG 360

CGCTCGTTCC CCCAAATGCT TGGGGAAGCT TGCTGAATCG CATTTACCAG ATGAGCCATG 420

TAATTAGAGT TATTAGAATA ACCTGCTCTA CGTGACAATG AATGATTATC ATCATGCCCA 480

ATCCAGCCAC CTAGGGTTAA TCTAGGTGTC GAAAGCATGA GCCACATATT TTCGTCTTGG 540

TTGGTTGTAC CAGTCTTCCC AATCCAATCT GCATTAGCCA GAGTAGGATT TAAAGAAGTC 600

AGGTTAGACT TGAAGGTTGT TGTCACACGA GAGGATAGAA CTTCTCGTAG CAATCCCTGC 660

ATAATCGTCG CAGTAGCTTT TGAATAGACT TGAACCGGTT TATCCTGATA CTCATACACC 720

ACTCTACCAT CTGCTGCTTC AATCTTTGAA ATCACATGCT TCTGATGATA AACTCCATTA 780

TTAGCTAAGG TCTGATAGCC ATTGGTATGC TGGGCAACTG TGACTTCAAT ACCACCACCC 840

ATTGGCAAGC TCTCAATACC GTACTCAGGA ATCTCGTAAC CCATCTTTTC CATATAACCC 900

TTGACATCAA CACCCTTTTC ACGGAGCATA CGATAGGTCC AGTAAGCAGG GATATTCCAT 960

GAATAGTTCA GAGCTTCTCC CAAGGTCATC ATTCCTGTTC CCTTGCTATT AGCATACATA 1020

ATCGGATTGC CATTAGCAAA GTTTGTTGGA TAGTTAGATA GAATCGTTTC ACTTCCCATC 1080

AAGCCCTGGT CAATAGCAAT ACCGTAGGCC AGCAAGGGCT TGGTAGTAGA AGCTGGCGAA 1140

CGTTTGGTAT CAAAGGCATG ATTATTTTGA TTTTCTTGAT AATTACGACC ACCTACAAAG 1200

CCTAGAATAG CACCTGTTTG GTTATCCATC AAGACATTCC CTACTTCTAC ACGACCTGTT 1260

CCATCGTCTA AAAGATAGCC ATAATCAGCA ACCGCACTTT GCATGGCAGA ATGAATTTTC 1320

TGATCTATGG TAGTAGTAAT CTTATAACCA CCATTTTCAA TTTCCTTGGC TGCCAAATCT 1380

CGATAAAACT TCTGAGTTGC CTCATTTTTC AACTCCTTAG CGGAGACATT GTCTCTCTGA 1440

GCTAGATAGT CATACATACG TTCTTGAGCT TCTGCCAAAG TTGTAAAGTA TAAATAGTCT 1500

CGTGAAATTC CTGTAACCGT GCCCGATGGT AAAAAGTCCT GTTTAAGGTC ATAATCCTTG 1560

TACTGAGAAT ACTCGTCTTT GCTTAATGCA CCTGTACGAT ACATACTGTA AAGAACTGCC 1620

TTAGCCCGTC TTAAGCCAAT TTCTAGGTCT TCATCACTCT TCAACTCCCC AGTATTTTCA 1680

TAAGGAGAGT AAGTAATGGG ACTCTGTGGA AGTCCTGCTA AAAATGCTGC TTGAGGAACA 1740

GTCAACTGAC TGGCATCTAC ACCGAAAATT CCCTCAGCTG CTTGCCGAGC CCCTGCAATA 1800

CTGCGGAATT TCAGGAGACA ATGACATTAG AAAGATGTCG ACTCTTTAGT CAAAAGTATC 3720

CAGAAAATAC ATTGATTGCG ATGGATGGTG TGAAGATAGT TGGTTTTATA AGTTATGGCA 3780

ACTGTCGTGA TGAGACTATT CAAGCTGGTG AAATTATTGC TTTATATGTT TTAAAAGACT 3840

ATTATGGAAA AGGAATCGCA CAAAAGTTAG TGAAAGCAGC TTTGACTGAT CTTAATCATT 3900

TTTCTGAAAT TTTCTTATGG GTATTGAAAG ATAACAAGCG CGCCATTGCT TTCTATCAAA 3960

AAATGGGTTT TACTTTTGAT GGACAAGAAA AAATACTTGA ACTTGGAAAG CCTATAAAGG 4020

AAAAACGGAT GGTATTCTAT TCTAAATAAT TCTCAAAAGT AAAAGCTAAT ATGGTACCAA 4080

GTCTGAAAAT TTAATAAATT AGAAAGCGAG TAAATTTATG TCCCGTTCCC AATTAACAAT 4140

TTTAACAAAT ATCTGTCTGA TTGAAGACCT CGAAACTCAG CGCGTGGTGA TGCAGTATCG 4200

CGCCCCTGAA AACAATCGCT GGTCTGGTTA TGCCTTTCCT GGAGGTCATG TAGAAAATGA 4260

TGAGGCTTTT GCGGAGTCTG TCATTCGTGA AATCTACGAA GAAACAGGGT TGACTATCCA 4320

AAATCCTCAA CTTGTCGGCA TTAAAAATTG GCCACTAGAT ACAGGTGGGC GCTATATTGT 4380

CATTTGTTAT AAGGCGACTG AGTTCTCTGG TACCCTTCAA TCTTCAGAAG AGGGAGAAGT 4440

TTCTTGGGTG CAAAAAGACC AGATTCCAAA CTTAAATCTG GCCTATGATA TGCTACCATT 4500

GATGGAAATG ATGGAAGCTC CCGACAAGTC AGAGTTTTTC TACCCTCGCC GTACAGAAGA 4560

CGATTGGGAA AAGAAAATCT TCTAGTCTTT TACTAAATAA CCTAGCTGAT CCAAGGCCTC 4620

CTCGATATAG TGGAGGTCTT GTTGTGTCTC GGCTTCAACT AGGTGATAAT GAATACCATC 4680

TGTTAACTCA GAAATTGGCT TAAAGTCAGA ACGTTCAACT TGTTCTAGAA AATGTTGCAC 4740

GTCGCGGCGA CAGGTCAGTT TTAGTAAGGT TTCAATCTCT CCATAAACAG GATGATCAAT 4800

CAAGATATTT TGAACGCGAC CACCATTATC TACGATAGCA AGTAATTCTC GTCCAATTTC 4860

TTCAACTTCA TGCTTGACCT TAAATAATTT GTGATGATAA GTATTTGCAT TAGCATCTTT 4920

ATAGATATAA CCACGATTGG TAGATAGAAT TGGAGATCCA TCAGCTCTTA AAATTGCAAT 4980

ATCTTGAACA ATAACTTGTC GAGTGACATG AAAGTGCTCA 5020

(2) INFORMATION FOR SEQ ID NO: 143:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4965 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

AAAAAGTGGC AATCCATTGA TTGGCCACTT CATTTAGAGA ATTATCGTCT CGCCCTTGAA 60

GTATGATATC GAAATGTTGC GTGAGATGGG CTATACCAAT GGGGTTGAAA ATTATTCTCG 1980

CCACATGGAT GGACGGAGCG AAGGAGAGCC TCCTTATACG CTTCTCGACT TCTTCCCAGA 2040

TGATTTCTTG ATTATGATTG ACGAGAGTCA TATGACCATA GGGCAAATCA AGGGCATGTA 2100

CAATGGAGAC CGTTCGCGTA AAGAAATGCT GGTTAATTAT GGTTTCCGTT TGCCGTCTGC 2160

TTTGGACAAT CGTCCTCTCC GTCGGGAGGA GTTTGAGAGT CACGTTCATC AGATTGTTTA 2220

CGTTTCAGCG ACACCTGGTG ACTATGAAAA TGAACAGACC GAGACAGTGA TTGAGCAAAT 2280

CATTCGTCCA ACGGGACTCT TGGATCCAGA GGTGGAAGTC CGTCCGACTA TGGGACAGAT 2340

TGATGACCTC TTGGGTGAAA TCAATGCCCG CGTTGAAAAA AATGAGCGTA CCTTTATCAC 2400

AACTTTGACC AAGAAAATGG CAGAGGATTT GACCGACTAC TTCAAGGAAA TGGGTATCAA 2460

GGTCAAGTAC ATGCACTCGG ATATCAAGAC CTTGGAACGG ACGGAGATTA TCCGTGACCT 2520

GCGCTTGGGT GTCTTTGATG TCTTGGTCGG AATTAACCTG CTCCGTGAAG GAATTGACGT 2580

TCCTGAAGTG AGCCTCGTAG CTATTCTCGA TGCTGACAAG GAAGGTTTCC TTCGCAACGA 2640

ACGTGGACTC ATCCAGACCA TTGGACGTGC TGCACGTAAT AGCGAAGGTC ATGTTATCAT 2700

GTATGCGGAC ACGGTTACCC AGTCTATGCA ACGTGCTATC GATGAAACTG CCCGCCGTCG 2760

CAAAATCCAG ATGGCCTATA ATGAAGAACA TGGTATCGTT CCACAAACCA TCAAGAAAGA 2820

AATCCGTGAC TTGATTGCTG TGACCAAGGC AGTTGCTAAG GAAGAAGACA AGGAAGTCGA 2880

TATCAATAGC CTCAACAAAC AAGAGCGCAA AGAACTAGTC AAAAAGCTTG AGAAACAAAT 2940

GCAAGAAGCA GTTGAAGTGC TTGACTTTGA ACTAGCAGCT CAGATTCGTG ATATGATGCT 3000

GGAAGTCAAG GCCTTGGATT AGGGGAATAG TATGATTTAT TTAAGAAAGT TAAAGAAAGA 3060

AGATTTGATG TCTTTATGGG AAATGGCTTA TTCACAGCTT AATCCAGTTT GGAAACAGTA 3120

TGATGCTCCC TATTATGATG ATTATCAGTA TTTTTCAAAT TTTAAAGAAT TCGAACTACA 3180

AAAATCAGAA TCCATTTTAA GCAACTCAAA TCGCCTTGGT ATTTTTGTTG ATGATAAACT 3240

AGTTGGGACT GTTTCGCGTT ATTGGGTATG TAAAGAAACA AGATGGATGG AATTGGGAAT 3300

TGGTATTTAT GATAAAAAAT TCTGGAACAC TGGTATTGGG AAAGTTGCTA TGTTGCAGTG 3360

GATAGATAGG ACGTTTCAGG ATTACTTGGA GTTGGAGCAT CTGGGTTTGA CAACTTGGTC 3420

AGGAAATATT GGTATGATGA AACTTGCTGA AAAATTAAGA ATGAAAAAAG AAGCTCATAT 3480

TCCAAAAGTT CGTTATTATC AAGGTAAATA TTTTGACAGT ATTAAATATG GTATTTTGAG 3540

AGAAGACTGG GAGAAAATAA ATGACGGTTA TTATCAAATC AATGGAAACT CCTGAAGAGA 3600

TAGAAGGTAA ATCCTTCGTT CACTGGCAAA CGTGGAGAGA GGCTTATGAT GACCTTTTGC 3660

ACTTTGTATG GTATTGGCTT GTTAGCTCTC ATCGGACTTG CTAGAAATTT TGGAGAGGCT 180

GGTCAAAATC TTGCAAGCTA CTTGCAGACC TTGCATCAGA GCTTGACGGA TAAAACAAGT 240

GACTTTCGTT TAATTTTAGG ATTACTGGCC TTTGGTTATT CTTAACACTG TGTTCAGATG 300

GACAAGAAAA GTTGAGAAAA GACCTATTCG AACCTTGGGA TTTTATAGAG AGAATTTCCT 360

CAGCAATCTT CTGAAAGGAT TTAGTCTAGG CCTGGCACTT TTTCTTCTGA CCTTGTTAGG 420

TTTAGTGGTC TTAGGTCAAT ATCGTTTGGA ATCCATTCAC TTGAATCCTT ATTCTCTTGC 480

CTTTGTCGTC TTTACTATCC CATTTTGGAT TTTACAGGGG ACAGCAGAAG AAGTGGTGGC 540

CCGTGCTTGG CTACTTCCTC AATTGGCCTC AAGAACCAAT CTAAAACTAG CTATTCTTAT 600

ATCTAGCCTG TTCTTTACCC TGCTTCATAT GGGCAATTCT GGTCTCACCC CTCTATCTCT 660

AGTAAATCTC TTTTTATTCG GAGTTGCCAT GGCTCTTTAC CTTCTCAAAA CTGATACAGT 720

TTGGGGTGTT GCAGGTATTC ATGGTGCTTG GAATTTTGCT CAGGGTAATC TCTTTGGGAT 780

TTTAGTTAGT GGTCAACCGT CAGaACGTCT CTGATGACCT TTTTACCACA AGGCAATCAA 840

GATTGGCTAT CAGGTGGTTC TTTTGGCATA GAAGGTTCCA TTATGACAAG TCTGGTATTA 900

CTACTGCTGA TTGTCTATCT TGCTAATAAA TTAAAGAAAG AAAATGAAAG GATGTGACTT 960

CGGTCCGTCC TTTTCTTCGT GAAAATACTA TAAGTATGCT AAAATAGGAA TAGCACATGG 1020

AGAGAGGATT CTTATGATCA ATCACATTAC AGATAATCAA TTTAAACTAG TATCAAAATA 1080

TCAACCATCA GGAGATCAAC CCCAAGCTAT CGAGCAGTTG GTGGATAACA TTGAGGGGGG 1140

AGAAAAAGCT CAGATTCTGA TGGGGGCGAC TGGAACAGGG AAGACCTATA CTATGAGTCA 1200

GGTCATTTCT AAAGTCAATA AACCAACTCT GGTTATTGCC CACAATAAAA CTCTGGCTGG 1260

TCAGCTCTAT GGGGAGTTTA AGGAATTTTT CCCTGAAAAT GCAGTTGAGT ATTTCGTATC 1320

CTACTATGAT TATTACCAGC CAGAGGCCTA TGTCCCTTCT AGCGATACCT ATATTGAGAA 1380

GGATAGTTCT GTCAATGACG AGATTGACAA GCTTCGCCAC TCAGCTACCT CAGCCCTTTT 1440

GGAGCGTAAT GATGTTATTG TCGTGGCCTC AGTCTCTTGT ATCTATGGTT TGGGTTCGCC 1500

CAAGGAATAC GCTGATAGTG TCGTTAGTCT CCGTCCTGGT CTAGAGATTT CTCGTGATAA 1560

ACTCTTGAAT GACTTGGTCG ATATTCAGTT TGAACGTAAT GATATTGATT TCCAACGCGG 1620

AAGATTTCGC GTTCGTGGGG ATGTGGTAGA GATTTTCCCA GCTTCCCGAG ATGAACATGC 1680

CTTTCGAGTA GAATTTTTTG GAGACGAAAT TGACCGTATT CGTGAAGTTG AGGCTCTGAC 1740

AGGTCAGGTG TTGGGAGAAG TGGATCATTT AGCGATTTTC CCAGCGACAC ACTTTGTGAC 1800

CAATGACGAC CACATGGAAG TTGCCATTGC AAAGATTCAG GCCGAGTTGG AAGAACAATT 1860

AGCTGTCTTT GAAAAGGAAG GTAAACTGCT TGAAGCCCAG CGTTTGAAAC AGCGGACAGA 1920

AAAATGTTTC CACGCTTCCA TCCTACCACT TGGAGGAAAG AGATTTCCAC ATTCCTCAAG 11640

TCAAGTGCTA TGAGTTAACA TTTGAAGGAA GCAAGGAAGG AAAGGTCTAT GCACGCATTG 11700

TTCTTCCAAA GAGTGAGGAG AAGGTCCCAT TAATCTTCCA TTTTCATGGT TATATGGGAC 11760

GTGGCTGGGA CTGGGCCGAC ATGCTGGGCT TCACCGTAGC TGGTTACGGT GTTGTTTCCA 11820

TGGATGTGCG GGGCCAGTCA GGTTACTCAC AAGACGGCTT GCGTTCTCCT TTAGGAAATA 11880

CCGTGAAGGG GCATATTATC CGTGGTGCTG TGGAAGGTCG GGACCACCTC TTTTATAAGG 11940

ATGTTTATCT GGATATTTAC CAGTTGGTCG AAATTGTTGC TAGTCTGTCT CAGGTTGATG 12000

AGAAGCGTCT TTCTAGCTAT GGTGCCTCAC AAGGAGGGGC TCTAGCTCTA GTTGCAGCAG 12060

CGCTCAATCC TCGAATTCAG AAAACAGTTG CCATTTATCC CTTCTTGTCA GACTTCAGAC 12120

GGGTGATTGA GATTGGTAAT ACTAGCGAGG CTTACGACGA ACTTTTCCGT TATTTCAAGT 12180

TTCACGACCC CTTCCATGAA ACAGAGGAGG AAATCATGGC GACCCTTGCC TATATCGATG 12240

TCAAAAATCT TGCCCATCGT ATCCAAGGTG AGGTTAAGAT GATTACGGGC TTGGACGACG 12300

ATGTTTGCTA TCCCATTACC CAGTTTGCGA TTTATAATCG TCTGACCTGC GATAAAACCT 12360

ATCGCATCAT GCCTGAGTAT GCTCACGAAG CCATGAATGT ATTTGTCAAT GACCAAGTCT 12420

ACAACTGGCT CTGTGGAAGT GAGATTCCTT TTAAATATCT AAAATAAGGA GTCGACTCTA 12480

AGCACAAAAT CTTAAAAATT ACAAACACGC ATAGTATCAG GGGATTAAGA AAACTTTATA 12540

CTATGCGTTT TATCATGGAA ATATAGTAAA ATGAAATAAG AACAGGACAA ATCGATCAGG 12600

ACAGTCAAAT CGATTTCTAA CAATGTTTTA GAAACAAATG TGTACTATTC TAGTGTCAAT 12660

CTATTATATT TATAGAATTT TTTGTTGCTA GATTTGTCAA ATTGCTTAAA ATAATTTTTT 12720

TCAGAAAGCA AAAGCCGATA CCTATCGAGT AGGGTAGTTC TTGCTATCGT CAGGCTTGTC 12780

TGTAGGTGTT AATACTTTTC AAAAATCTCT TCAAACCACG TCAGCTTCGC CTTGC 12835

(2) INFORMATION FOR SEQ ID NO: 142:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5020 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

GGGGATATGA AGAACAAAAG AATATTTAAA GACTTCCAAG CTTCAAAAAT GAGTTTAAAC 60

ATTTACACAA GCCCCTTGTT AGCCTTTGTT TTTGTCTTCA TAGGAGAGTT TGTGGCTTTT 120

TCGTCTCCTA CAAGGGGATG TGGGGAGTGG AAAAACGGTA GTCGCTGGCT TGGCCATGTT 9840

TGCGGCAGTG ACAGCAGGTT ATCAGGCTGC CCTAATGGTA CCAACAGAAA TCCTCGCAGA 9900

GCAACACTTT GAGAGTTTAC AGAACCTTTT TCCCAATTTG AAACTGGCTC TCTTGACAGG 9960

TTCCTTGAAA GCTGCAGAAA AGAGAGAAGT CTTGGAGACC ATTGCCAAGG GTGAGGCTGA 10020

TTTGATTATA GGAACTCACG CTCTGATACA AGATGGGGTG GAGTATGCTC GTCTTGGTTT 10080

GATTATTATC GATGAGCAGC ACCGTTTTGG TGTAGGGCAA AGGCGTATTT TACGGGAAAA 10140

AGGTGACAAT CCAGATGTCC TCATGATGAC GGCGACTCCC ATTCCACGGA CGCTTGCCAT 10200

CACAGCCTTT GGAGATATGG ATGTTTCCAT TATCGACCAG ATGCCAGCAG GTCGGAAGCC 10260

TATTGTGACG CGCTGGATCA AACATGAGCA ACTACCTCAG GTCTTGACTT GGTTAGAGGG 10320

GGAAATTCAA AAAGGTTCCC AAGTCTATGT CATCTCTCCT TTGATTGAAG AATCAGAAGC 10380

TCTAGATTTG AAAAATGCCA TTGCCTTATC AGAGGAGTTG ACGACTCATT TTGCAGGCAA 10440

GGCAGAGGTG GCTCTTCTAC ATGGTAGGAT GAAGAGTGAC GAAAAAGACC AGATCATGCA 10500

GGATTTCAAG GAGAGAAAGA CGGATATTCT GGTTTCGACG ACGGTTATTG AGGTTGGGGT 10560

CAACGTTCCC AATGCGACTG TCATGATTAT CATGGATGCC GATCGCTTCG GTCTCAGTCA 10620

ACTTCACCAG CTTAGAGGTC GTGTCGGTCG GGGGGACAAG CAGTCCTACG CTGTTCTCGT 10680

TGCTAATCCC AAGACGGATT CTGGGAAAGA CCGCATGCGT ATCATGACAG AAACGACCAA 10740

TGGATTTGTC CTTGCGGAGG AAGATTTGAA AATGCGTGGT TCTGGTGAGA TTTTTGGAAC 10800

CAGACAGTCA GGACTTCCAG AGTTCCAAGT GGCTGATATT ATCGAAGATT TTCCGATTTT 10860

AGAAGAAGCA AGAAAGGTTG CTAGCTACAT TAGTTCTATA GAAGCTTGGC AAGAAGATCC 10920

AGAGTGGCGC ATGATTGCCC TTCATCTGGA AAAGAAAGAA CATCTGGATT AAGCTTTCTC 10980

TAAGGAAAAC TTATACTCAA TGAAAATCAA AGAGCAAACT AGGAAGCTAA CCGCAGGTTG 11040

CTCAAAACAC TGTTTTGAGG TTGTGGATGA AACTGACGAA GTCAGCTCAA AACACCGTTT 11100

TGAGGTGGCA GATAGAACTG ACGAAGTCAG TAACATATAT ATACGGTAAG GCGACGCTGA 11160

CGTGGTTTGA AGAGATTTTC GAAGAGTATT AAGCTAGTTT TTAGGTTTGG CTCTTATACT 11220

AGAGTCATCA AAAAGAAACG AGGACTCTCA TATGACAGTA ACGATTAAAG TAAATTACCA 11280

AACCACTTTC CAAAAGAAGG AAGCAAAAAA CTAGTATAAA CAGAAGAGAG AGCGAAATGC 11340

TCTTTTTTCG TTTCTAAAAC TACTTTCAGC CCATCATCCT AAAAGTAAAG AATCTAAATT 11400

CACTTTCTAT TTACCCTTCT TTCTTGCATT GATTACATAG ATATGCTACA GTTGTGGTAA 11460

CGATTACAAA ATAAAAGGAG CATGCTATGA AAAATCCAGC TTTGCTAGAA GAAATTAAGA 11520

CCTATAGAGG AAGGGATGAG GTTCCGGAAG ACTTTGATGA TTTCTGGGAT GGGGAAGTGA 11580

AGCCATTGAA CTCAGACAAG CTGGACTCAG CAAGCCAATC CTCATTTTAG GAGTTTCTGA 8100

AATCGAAGCT GTTGCTCTAG CTAAAGAATA TGACTTCACC TTGACAGTGG CTGGACTGGA 8160

GTGGATTCAA GCACTCTTAG ATAAGGAAGT GGACCTAACT GGATTGACAG TCCACCTCAA 8220

GATTGATTCA GGGATGGGAC GGATTGGTTT TAGAGAGGCA AGTGAGGTTG AGCAGGCTCA 8280

AGATTTGCTC CAACAACACG GTGTTTGTGT TGAAGGAATC TTTACCCACT TTGCTACTGC 8340

TGATGAGGAA TCAGATGACT ATTTTAATGC CCAGTTAGAA CGGTTTAAAA CTATTTTAGC 8400

TAGTATGAAG GAAGTTCCAG AGCTGGTTCA TGCTAGCAAT TCTGCAACGA CTCTTTGGCA 8460

TGTAGAGACT ATTTTCAATG CGGTTCGTAT GGGAGATGCC ATGTATGGCC TCAATCCAAG 8520

TGGAGCGGTC TTGGATTTGC CTTATGATTT GATACCGGCC TTGACCTTGG AGTCTGCTCT 8580

GGTTCATGTC AAGACAGTTC CAGCTGGAGC TTGCATGGGC TATGGAGCAA CTTATCAAGC 8640

GGATAGCGAG CAAGTCATCG CGACCGTGCC AATCGGGTAT GCAGATGGAT GGACAAGAGA 8700

CATGCAAAAT TTCTCTGTCT TGGTAGATGG CCAAGCTTGC CCAATTGTCG GCAGGGTTTC 8760

GATGGACCAA ATCACTATTC GATTGCCTAA GCTTTATCCG CTAGGAACCA AGGTAACCTT 8820

GATTGGCTCC AATGGGGATA AGGAAATCAC TGCAACTCAG GTAGCGACCT ACCGCGTAAC 8880

CATTAACTAT GAGGTGGTTT GCCTCCTCAG CGACCGTATT CCGAGAGAAT ATTATTAGAA 8940

AAGAAAGGAG TGGAGCATGA ATCTACATCA ACCCTTGCAT GTCTTGCCTG GTGTGGGACC 9000

AAAGTCAGCA GAAAAATACG CCAAACTAGG AATTGAAAAC TTGCAAGATC TCTTGCTCTA 9060

CTTTCCTTTC CGTTATGAAG ACTTCAAAAC CAAGCAGGTG CTGGAGCTGG AAGACGGTGA 9120

GAAGGCAGTT CTTTCTGGTC AGGTAGTGAC TCCTGCTAGT GTCCAGTATT ATGGTTTCAA 9180

GCGCAATCGC CTGCGTTTTA GTCTCAAGCA GGGAGAGGTC GTTTTTGCGG TGAATTTCTT 9240

TAACCAGCCC TATCTGGCTG ATAAAATAGA GTTGGGAGCA ACCCTTGCTG TCTTTGGAAA 9300

ATGGGACCGC GCTAAGGCTA GTCTGACTGG GATGAAGGTT CTGGCTCAGG TAGAAGATGA 9360

CCTCCAGCCT GTCTATCGTC TGGCTCAGGG AATCAGTCAG GCCAGTCTGG TCAAGGTCAT 9420

CAAGACGGCT TTTGATCAGG GACTGGACCT CTTGATAGAA GAAAATCTGC CCCAGTCTTT 9480

ACTAGACAAA TACAAACTCA TGTCCCGTTG TCAGGCAGTC CGTGCTATGC ATTTTCCAAA 9540

GTATTTGGCA GAATACAAGC AGGCTCTTCG CCGTATAAAG TTTGAGGAAC TCTTTTATTT 9600

CCAAATGCAG CTGCAGATGC TCAAGTCTGA AAATAGAGTT CAGGGAAGTG GTCTGGTTCT 9660

GAATTGGTCT CAGGAAAAAG TGACAGCAGT TAAAGTAAGT CTTCCTTTTG CCCTGACCCA 9720

AGCTCAGGAA AAGAGTTTGC AGGAAATTTT AACTGATATG AAGTCCGACC ACCACATGAA 9780

GAAAAAATTA AGCAGTTCGT TCGTGGTTTT ATGATTGAGT CTTATCTGGA AGATGGTCGA 6300

CAAAATGAGC CAGAAGTATT TGGTAAGTCT ATCACAGACC CTTGCCTGGG TTGGGATAAC 6360

ACAGAAGCTC TTGTCAGAGA AATTTACAAA ACGTTAGGAG AATAAGATGG CATTTATTGA 6420

AAAAGGTCAA GAAATCGATA TGGAAGTCAT CAAGGCTGAA ACCCAATTGT CTGCGGAAGC 6480

CTTGAGACTC AAGGAAAGCC GTGACAGGGA ATTGGCAGAT ATTATTTCAG GGGAAGATGA 6540

CCGTATTCTC TTGGTGATTG GTCCTTGCTC TTCTGATAAT GAAGAGGCGG TCTTGGAATA 6600

TGCTCGCCGT TTATCTGCCT TGCAAAAGAA GGTAGCGGAT AAGATTTTCA TGGTCATGCG 6660

CGTGTATACT GCTAAGCCTC GTACCAATGG AGACGGCTAT AAAGGATTAG TTCACCAGCC 6720

AGATACTTCT AAGGCTCCAA GCCTGATTAA TGGCTTGCAG GCTGTGCGCC AGTTGCACTA 6780

CCGCGTGATT ACAGAGACTG GTTTGACAAC GGCAGATGAG ATGCTTTATC CGTCAAATCT 6840

GATCTTGGTG GATGACTTGG TCAGCTACCA TGCCGTTGGA GCTCGTTCTG TGGAAGACCA 6900

AGAGCACCGC TTTGTGGCTT CTGGGATTGA TGCACCAGTA GGGATGAAAA ATCCAACCTC 6960

AGGAAATTTG GGTGTTATGT TTAACGCCAT CTATGCTGCT CAAAACAAGC AAACCTTCCT 7020

TTATCATGGG CAGGAAGTTG AGACATCAGG TAATCCTTTG GCCCATGTTA TCCTCCGTGG 7080

AGCAGTCAAC GAGTATGGCA ATTATATGCC GAATTACTAC TATGAAAATC TACTCCAAGC 7140

CATTGAACGC TATGAAACCA TGGGACTTGA AAATCCTTTT ATCCTCATTG ACACCAACCA 7200

TGATAACTCA GGCAAGCAAT ATATGGAGCA GATTCGAATT GTTCGCCAGA CCTTGCAGAA 7260

TCGTGATTGG AATGAGAAAA TTAAAAAGAC GGTTCGAGGA TTTATGATTG AATCTTACCT 7320

AGCAGATGGT CGTCAAAACC AACCAGAGAT CTTTGGTTGC TCTATTACTG ACCCTTGCCT 7380

AGGTTGGGAA AATACAGAGG CCTTGGTAGA AGAGATTTAT GTTACCTTGA CAAAATAAGT 7440

GAAAAGGATG GAGTTGGGGA ATCTCAACTC CTTTTGATGA GAATGATAGT TGGACACGGA 7500

ATTGACATCG AAGAATTGGC TTCGATAGAA AGCGCAGTTA CACGACATGA AGGATTTGCT 7560

AAGCGTGTAC TGACCGCTCA GGAAATGGAG CGCTTCACCA GTCTCAAAGG ACGCAGGCAA 7620

ATAGAATATT TAGCTGGTCG CTGGTCGGCT AAGGAGGCCT TTTCCAAGGC TATGGGAACG 7680

GGCATTAGCA AGCTCGGTTT TCAGGATTTG GAAGTCTTGA ACAATGAACG TGGGGCGCCT 7740

TATTTTAGTC AGGCACCATT TTCAGGAAAG ATTTGGCTGT CTATCAGCCA CACCGATCAG 7800

TTTGTGACAG CCAGTGTCAT TTTGGAGGAA AATCATGAAA GCTAGTCCAC ATAGACCAAC 7860

CAAGGCTCTG ATTCATCTGG GAGCTATTCG ACAAAATATT CAGCAAATGG GGGCTCATAT 7920

CCCTCAAGGA ACGCTCAAGT TGGCTGTGGT TAAGGCCAAT GCTTATGGTC ATGGAGCTGT 7980

TGCCGTTGCC AAGGCAATTC AAGATGATGT TGATGGCTTT TGCGTTTCCA ATATCGATGA 8040

GACGCGTCAG GTTGAAGCAG CTCAGAAACG TGTCGAAGGA AATAACTACG ATACCCGTAA 4560

ACAAGTCCTT CAATACGATG ATGTCATGCG TGAACAACGT GAGATTATCT ATGCTCAACG 4620

TTACGATGTC ATCACTGCAG ATCGTGACTT GGCACCTGAA ATTCAGTCTA TGATCAAACG 4680

CACGATTGAA CGTGTCGTTG ATGGTCATGC GCGTGCCAAA CAAGATGAAA AACTAGAGGC 4740

AATTTTGAAC TTTGCTAAGT ACAACTTGCT TCCTGAAGAT TCTATTACGA TGGAAGACTT 4800

GTCAGGCTTG TCTGATAAGG CCATCAAGGA AGAGCTTTTC CAACGTTCCT TGAAGGTTTA 4860

CGATAGTCAG GTTTCAAAAC TACGCGATGA AGAAGCAGTT AAAGAATTCC AAAAAGTTTT 4920

GATTCTACGA GTGGTGGATA ACAAGTGGAC AGATCATATC GATGCCCTTG ATCAATTGCG 4980

TAACGCGGTT GGACTTCGTG GCTATGCTCA GAACAACCCT GTTGTTGAGT ATCAGGCAGA 5040

AGGTTTCCGT ATGTTTAATG ATATGATTGG TTCGATTGAG TTTGATGTGA CACGCTTGAT 5100

GATGAAAGCA CAAATTCATG AACAAGAAAG ACCACAGGCA GAACGTCATA TCAGTACAAC 5160

AGCGACTCGC AATATCGCTG CTCACCAAGC AAGTATGCCA GAAGATTTGG ATTTGAGCCA 5220

GATTGGACGC AATGAACTTT GCCCATGTGG TTCTGGTAAG AAATTTAAAA ACTGTCACGG 5280

TAAAAGACAA TAAAATGAGA TAGTTTAGAG GCGGATATCT TGTGAAAAGT AAATTTTTAC 5340

TGGGTATCCG TTTGCTTTAT AAGGAGATGA GTTATGGTAT TTACAGCAAA AAGCTCTAAA 5400

ATAAATATAG AAGAAGTTCG TGCCTTGTCA AAATTAGAAG GTCAGGCTTT GGAGAGGAAA 5460

TCACAGCGAG ATCAAGAGCT AGAAGCCATT ATACGTGGAG AAGACCAGCG AATTCTCTTG 5520

GTAATCGGGC CATGCTCATC TGACAACGAA GAAGCTGTCC TTGAATACGC TAAGCGTTTG 5580

GCAGTCCTAC AAGAAGAAGT GGCAGATCGT ATCTTTATGG TTATGCGTGT TTATACTGCC 5640

AAACCCCGTA CCAACGGAGA TGGCTATAAG GGCTTGATTC ACCAGCCTAA CGCGACAGAA 5700

GCGCCTAGTC TTATCAATGG AATCAAAGCC GTTCGCCATC TTCACTATCG TGTCATCACA 5760

GAAACAGGGA TGACAACTGC TGATGAAATG CTTTATCCTG AAAACCTTCC GCTTGTACAT 5820

GATTTGATTT CTTACATGGC AGTTGGTGCC CGTTCAGTTG AAGACCAGCA ACACCGCTTT 5880

GTGGCAAGTG GGGCAGGATT TTCTACTGGT TTTAAAAATC CAACCTCTGG AAATCTCAAT 5940

GTCATGTTTA ATGGGATTTA TGCTGCTCAA AACAAACAAA GTTTCCTTTT CTTAGGAAAA 6000

GAAGTAGAAA CAACTGGGAA CCCGCTTTCA CACGCTATTC TTCGTGGTGC TCTTAATGAG 6060

TATGGAAAAA ATATTCCCAA CTACTATTAT GACAATTTAA TTGATACCAT TGCCCAGTAT 6120

GAGAAAATGG GCTTGGAAAA TCCTTTTATC ATCATTGATA CCAATCATGA CAATTCTGGT 6180

AAGCAGTATA TTGAACAGAT CCGAATTGTC CGCCAGACCT TGATTAACCG TGCTTGGAAT 6240

CTTCCTTGAA AATAATCAGT AATTGTGCTA AAATTAAAGG AACATTCTAA AATATTCGGA 2760

ATTTAAAGTA AGGAAAAACA TGGCTAATAT TTTAAAAACA ATTATCGAAA ATGATAAAGG 2820

AGAAATCCGT CGTCTGGAAA AGATGGCTGA CAAGGTTTTC AAATACGAAG ACCAAATGGC 2880

TGCTTTGACT GACGACCAAC TAAAAGCAAA AACAGTTGAA TTTAAGGAAC GTTATCAAAA 2940

TGGAGAATCA CTGGATTCAT TGCTTTACGA AGCATTTGCG GTTGTCCGTG AAGGTGCCAA 3000

ACGTGTCCTA GGTCTCTTCC CTTATAAGGT TCAGGTCATG GGGGGGATTG TTCTTCACCA 3060

TGGTGACGTG CCAGAGATGC GTACAGGGGA AGGGAAAACC TTGACTGCGA CCATGCCGGT 3120

ATACCTCAAT GCCCTTTCAG GTAAAGGGGT TCACGTAGTT ACGGTTAATG AATACCTGTC 3180

AGAACGTGAC GCGACTGAGA TGGGTGAATT GTACTCTTGG CTTGGTTTGT CAGTAGGGAT 3240

TAACTTGGCT ACCAAATCTC CAATGGAGAA AAAAGAAGCC TATGAGTGTG ATATTACTTA 3300

CTCAACTAAC TCAGAAATCG GATTTGACTA CCTTCGTGAC AACATGGTCG TTCGCGCCGA 3360

AAACATGGTA CAACGTCCGC TTAACTATGC CTTGGTCGAT GAGGTTGACT CTATCTTGAT 3420

TGACGAGGCT CGTACACCTT TGATTGTATC AGGTGCCAAT GCGGTTGAAA CCAGTCAGTT 3480

GTATCACATG GCAGACCACT ATGTAAAATC TTTGAACAAA GATGACTACA TCATCGATGT 3540

GCAGTCTAAG ACTATTGGTT TGTCTGATTC AGGGATTGAC AGGGCTGAAA GCTACTTCAA 3600

ACTTGAAAAC CTCTATGACA TCGAAAACGT GGCTTTGACT CACTTTATCG ATAACGCCCT 3660

TCGTGCCAAC TACATCATGC TTCTCGATAT TGACTATGTG GTGAGCGAAG AGCAAGAAAT 3720

CTTGATTGTC GACCAATTTA CAGGTCGTAC CATGGAAGGT CGTCGTTATT CTGATGGATT 3780

GCACCAAGCT ATTGAAGCCA AAGAAGGTGT GCCAATCCAG GATGAAACCA AGACATCTGC 3840

CTCAATCACG TACCAAAACC TCTTCCGTAT GTACAAGAAA TTGTCTGGTA TGACGGGTAC 3900

AGGTAAGACT GAGGAAGAAG AATTCCGTGA AATCTACAAC ATTCGTGTTA TTCCAATCCC 3960

AACAAACCGT CCTGTTCAAC GTATTGACCA CTCAGACCTT CTTTATGCAA GTATCGAATC 4020

TAAGTTTAAA GCGGTTGTCG AAGACGTTAA GGCTCGTTAC CAAAAGGGTC AACCTGTCTT 4080

GGTTGGTACA GTAGCGGTTG AAACTAGTGA CTACATTTCT AAGAAATTGG TTGCAGCTGG 4140

TGTTCCTCAC GAAGTCTTGA ATGCCAAAAA CCACTATAGA GAAGCCCAAA TCATCATGAA 4200

TGCTGGTCAA CGTGGTGCCG TTACCATCGC AACCAACATG GCGGGTCGTG GTACCGACAT 4260

CAAGCTTGGT GAAGGTGTTC GTGAACTTGG AGGACTTTGT GTTATTGGTA CAGAACGTCA 4320

TGAAAGTCGT CGTATCGATA ACCAGCTTCG TGGACGTTCA GGTCGTCAAG GAGATCCAGG 4380

TGAGTCACAA TTCTACCTAT CTCTTGAAGA TGATTTGATG AAACGTTTTG GTTCTGAACG 4440

CTTGAAGGGA ATCTTTGAAC GCTTGAACAT GTCTGAAGAG GCCATTGAGT CTCGCATGTT 4500

GGGATGTAAG TTTGATGAAA CTCTTTACTC CTTATATCCT TTATATTGTC CCTTACATGG 1020

TGCTTGAAAA ATATGAAGAT AATGTTTAAG AATTTTAACA ATATTTTGCT AAATAGAAAG 1080

ATTGTTTTAC TACTTCGTAT AGTTCTGATG ATGATTTTGA TAAACCATCT ATTGTCAACA 1140

GCGGTTCAAA AGCAGGATGC TGTTATCTTT TTCAAGAGAG AATTGATTTC AATTTTTTCC 1200

TATAATGACT ATTCTGAAGC GAATTTAGAA ATCCCCAAAC TATTGTTAAA CCTTTCGCTT 1260

TTCATGGTAG GATGGCTCTC TGTCATTTTA CTTGAAAGTG ATTTGGCAGA CCATTACCAT 1320

CACTTGATTC GCTATCAATC AAGCTCCTTT TTCGATTATA CAAGGAAACG ATTGGTTGTC 1380

ATTTCTAAAT TTTTTACTCA AGATTTGTTT GTCTGGTTTC TTGGTTTACT TCCTCTAGGA 1440

ATTCATTTCA AAACAGTCGC ACTTTTCTTT TTACTTGCTC AGTTAATGAT GTTGTACTTA 1500

CTACTGTCTT ATCTGATAGC ACTGATTAGT GCGGGCGCTG GTTTTTCCTT TTTTCTCTAT 1560

TTTTTAGCAT TTGTGGGACA AGAATGGATG ATGGATCATA TTGTAACAGT GTATTTAGTA 1620

CTCTTAAGTT TATTAGTTAT GTTGATTGTT AGTCGCTTGG AAGAGAAATT TAAGAAAGGA 1680

TAAACGATGA GACTTGAAAT TATAAATGGA CAGAAAATTT ATGGGAAAAG ACCTATTTTA 1740

AATCAGTTGA ATTTGGTGTT TCAATCAGGA AAAATTTATG GACTTAAAGG TGATAATGGA 1800

TCTGGCAAGA CGGTTCTTTT AAAGATACTT GCTGGTTATA TTAAGCTTGA CAAAGGAAAA 1860

GTTCTTCAAG ATGGTAAAGT TTACGGGGTA AAAAATCATT ATATTCAGGA TGCAGGAATT 1920

TTAATTGAAA AAGTCGAGTT TTTATCTCAT TTATCCCTGA GAGAAAATTT GGAACTGTTA 1980

AGGTATTTTT CATCTAAAGT TACGGAAAAA AGAATTGCCT ATTGGATTCA ATACTATGAT 2040

TTACAGGAAT TTGAAGACAT TGAATACCGT CATTTATCCT TAGGAACAAA GCAAAAAATG 2100

GCCTTGATTC AAGCCTTTAT TTCCTCTCCT TCTATACTCT TTCTCGATGA ACCTATGAAT 2160

GCTTTGGATG AGAAGAGTGT GAGGTTAACC AAACAGGTCA TTTTATCTTA CCTGAAAAAA 2220

GAAAATGGTC TGGTTATCCT GACGTCGCAC ATATCGGAAG ATATTTCAGA CCTTTGTACA 2280

GATGTATTAG TTGTCGAAAA TGGACATATA CAAATGTAAA GGATATACAA TCCTAGGAGA 2340

TGGCTTATGG CACATCTAAA ATCATTTATT ACACGATATT CCAAGGTTTA TATTGGTTTA 2400

GTTCTGCTGA TCTGGCTGTC TTTCTTCTTT ATCCCTTGGG ATAAACCACT TCTGGGGATA 2460

AGGATTGACA TCTTCATCAT ACAGAAAATC TTGCTAGCTT TTGGAATTCT GTCCATTCTC 2520

ATGGCCTTGC TGTCCAAGAA AGTCAGTCTC TTTGTTTTTG GACTGATTTG CTGTCTTTCT 2580

CTTTGGATTA ACTTATTTAT CACATTTGCC ATTTTGCCGA TTTTTGGCAA TTAAACAGTC 2640

ATAAAAGTCG GAGAGGTTAG CTTGAAAACT AACCTCTTTT TCCTTTTCAA AATGGGGATT 2700

TTCAAAATTG CCAAGTTTAA CGAATTCTTT ATAATAATTA GCTACATCAT TACCAAGTAT 28500

AATTTGATAT TGTCCATTCT TTTTCATAAT ACCTATTACA CCTGGTATCT TCTTCACATC 28560

ATCATCATTG ACTAAATTTT CATCTTTTAA TTCTAATCTT AAACGTGTTA CACAATGGGT 28620

AACTCTATTG ACATTTTTTT CACCTCCAAT TACATCGAGG ATTTTTTGTA CCGTATCTTT 28680

ATAACTCATG GTATTCTCCT ATTCTATTAA TCTAAATTTT TTGTTAAGCG ACGAATATGA 28740

GCCATCAAAT AAACTAATTC ACTAGAAGTC AGCAAATAAT TGTACTCCGT TTGTATAAAC 28800

ATTGCTACCT GTTCACCACA TTCATATTCT CTAGGATATT TATTTTTCAT TAATGCTAAC 28860

AAGTCTTCAT CATCATCGTC GG 28882

(2) INFORMATION FOR SEQ ID NO: 141:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12835 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

GCCTATGTCT TTTTCAAAAA AATGCTTGAC TTGAGACGGG AACTAGGGAA GTCTAAAGGC 60

GGAAGGCATT GATTTATACT CTTCGAAAAT CTCTTCAAAC CACGTCAACG TCGCCTTGGA 120

TTATATATGT AACTGACTTC GTCGATGCTT ATCTACAACC TCAAAGCAGT GCTTTGAGCA 180

ACTTGCGGCT AGTTTCCTAG TTTGCTCTTT GATTTTCATT GAGTATTATA TTACTTTCTA 240

TTTGTAGGAG GTGGCTTATG AAGATTCCTC TCTTAACTTT TGCAAGGCAT AAATTTGTTT 300

ATGTCTTGCT TACTTTGCTT TTTCTTGCTT TGGTTTATCG TGATGTTTTG ATGACTTATT 360

TCTTTTTTGA TATTCATGCG CCCGATCTAG CTAAATTCGA TGGACAAGCA ATTAAAAATG 420

ACTTATTAAA ATCAGCATTA GATTTTCGTA TTCTCCAGTT CAATCTAGGT TTTTATCAAT 480

CATTTATTAT TCCAATCATC ATTGTTTTGC TAGGTTTTCA ATATATTGAG CTGAAAAATA 540

AAGTTTTACG ATTGAGTATT GGAAGAGAAG TGAGTTATCA AGGGTTAAAA AGAAAGTTGA 600

CTTTGCAAGT TGCAAGTATC CCTTGTTTGA TATATTTAGT GACTGTGCTG ATAATTGCAA 660

TTATAACCTA TTTCTTTGGG ACTTTTTCTC CTCTTGGATG GAATTCTCTA TTTTCTGATG 720

GAAGTGGTTT ACAAAGACTC CTAGATGGAG AGATAAAAAG CTATTTGTTC TTTACTTGTG 780

TCCTACTAAT CGGTATTTTC ATCAATGCAA TCTATTTTTT ACAAATAGTT GATTATGTGG 840

GGAATGTGAC TCGTTCGGCA ATCACCTATT TGATGTTTCT TTGGCTTGGT TCTATGCTGC 900

TTTATAGTGC CTTGCCTTAC TATATGGTTC CTATGACGAG TTTGATGCAA GCTAGCTATG 960

TTCTGTTGCT TCTCCTAATC CACCTTTGGG TAACACATCC TGAACTGATA AGCCCTTACC 26760

ATCTTCATTA TATGCTCCCT CTACTTGATT AGCTGCAACA GCTCCACCCC AAAGAAAATC 26820

ATCTGGAAAA ATGGTCATAA CTTTCCTCCA TTATAATATT ACCAGTAATT CCTTAGAATG 26880

CTCGATTGTC TGATTATTAG GTAATACTAA TACATCTAGA AAATCATTGG TATTCGTTAC 26940

AATTACTGGT GTAACTGTTT CGTAGCCTTT AGTCTTGATT AAATTCAAGT CCATTTCAAA 27000

AATCAACTGA TTTTTGAAAA CTCTGTCTCC TTCTTCTACA TGACTAATAA AACCTTGACC 27060

TTTTAGCTCA ACAGTATCTA ATCCAATATG AATTAGTAAC TCAACACCCT CATCACTCTT 27120

CAATCCAATT GCGTGCTTAG TCGGAAAAAT ATTTGTAATT TTCCCATCAA ATGGTGCATA 27180

AACCTTACCT TCACTTGGGA TAATCGCTAC TCCGTCTCCA ATTAGTTTAT CTGAAAATGT 27240

TTTATCCTGG ACATCGCTTA ACGGAATGAT TTCTCCTGAT ATAGGAGAAA ATATCATTTT 27300

TTTATTTGAA ACTCCAGCTT CAACTTCTAA ATTGCTAGAA CTCTCTTCTT CATCGATTCC 27360

AAATATATAA GCTAATACAA AGGTAATAAC AACCGAAATG ACCGCCACAA TTAAAGCATT 27420

TACAATATTT GATGGCACAT CAGAATAAAT AAATTGAGGC AACGCTATCA AAGATGGGAC 27480

AGCAAATAGA TATGCTTTAA CACTAGTAAG ACCTGCAAAT AATCCCGCTA ATCCACCACC 27540

AATCATAGCT GCATAAAGCG GTTTTTTATA TTTTAAAGTC ACACCATATA ATGCAGGTTC 27600

GGTAATCCCT GCAAGTAAGG CTGAGAAACC TGCTGCAAAA GCAATTTGTT TTGTATTATT 27660

ATTTTTACTC TTTAATGCAA CAGCCATCGA AGCAGCCCCT TGAGCTAAGT TTGACCCTAA 27720

CATTGCTGGA AGAATTAATA CGTCTGGAGT AGCAATAGAT GCCGCCAAAA AAATAGGTGC 27780

AAAAGCCCAA TGCATTCCAG TCATAACAAT AAATGGCATA ATAGCACCAA GAATAGCTAA 27840

TGTAAGCCAT CCAGCTACAC CATACATTTG CCCAACTAGA TTTGATAATC CTTCACCAAC 27900

AATTACTCCA ATAGGTCCGA CTACAACTAA GGCAATACAG CTTGATACTA ATAATACTAG 27960

CGTAGGTTGC AAAAAACTCT TAGTAATAGC TAGTGTTAAT TTAGCAATTA TTTTTTCAAT 28020

ATATTTCATC AACCAAACCA TAATAAGAAT TGGAACGACT GATGAACCAT AACTAGCTGG 28080

TGTCACAGGT GCACCAAATA AACTAAGAGG ATTCCCTGAT TGCACCATTT GAACAAAATT 28140

TGGATGGAGA AGTACACCTG CTACAGACAT AGCTAATGTA GATGTTACTT TTAATTTTTG 28200

TGATGCAGAA TAAGCTAATA ACAGCGGTAA GAAATAATAT GGAGCATCCC CAAAAAATGT 28260

CAAAAAAGCA ATAGTCTGAG AATCTGATTG CAATATACCA AGCATTGGTA AAATGATTAC 28320

CAAGACTTTC AACATACCTC CCCCTAACAT TGCTGGAATG ATTGGAGTCA TGGAACCAGC 28380

GATATACTCA ATGATTCTTT CTAAAATATT CCCTTTGTGC CCTTGAACAA CTGAATCGGA 28440

TTTCTTCGCG AAGCGCTTTT AATTGTTCTT CAATAGTTGA CATATTTCCT CCATCAGTCT 24960

CTCGTAGATA AAAAGAAAAC CACATGCCAA AAACTCCACT CGGAGCGTTG ACACGCGGTA 25020

CCATCCGTTT TCATCTGACA AGTCAGACCT TCATTTCTAA ATCCATGCGC AAGTGAATTC 25080

ACCCAGCTTT CATATAGAGA GCTTGCAGTC ACGGCTCTCC TCCCTGATAT ACTTCCCTTG 25140

GGCTACTAGT CTTTCAGATT CCTATTCAAT TACTACTTAG TTTATCAGAT TTTTACCATT 25200

CTTGCAAGAC CTATCTTACT TCTGCTTGTT AGCTTATTCT TATCTAAATT TATATAAACc 25260

TTATCTAAAT TAACTATTTA TAATTTTTGT AACAAAATTA AATTAATTGA CACTCCCCTA 25320

TAAAATAAAG AAGTTTAGAA TTTAATGTCT TCCAAACTTC TTTATTCCAT ATTTAATGAA 25380

ATGCCACCTT AACCGTGATA ATAGCTAGTC ATCAATAAAA AACTATTTGA ATAAGGATTC 25440

TCCATTTGAT TCAATCACTT CTTTATACCA AGTAAAAGAC ATTTTCTTAT ATCGATTTAA 25500

TGTACCACTT CCATCATCGT TTCGATCAAC ATAAATGAGA CCGTACCTTT TAGAAAGTTG 25560

TGCAGTGGAC ATAGAAACAC AGTCAATACA TCCCCAAGAC GTATAGCCCA TAATTTCAAC 25620

ACCATCCTGT AGAGCTTCAG CAACTTGCAA TAAATGTTCT TTCATATACT GAATTCTATA 25680

ATCATCTTGG ACGGTTAAGT TATTAAGTTC ATCTTTTATT AGTTGATCTT TAGCACCTAA 25740

TCCATTTTCT ACTATAAATA ATGGGATTTG ATAACGGTCA TAATATCTAT TTAAAATTAT 25800

ACGTAGTCCA ATTGGATCAA TTTGCCATCC CCACTCTGAA GACTCTAAAT AAGGATTTAC 25860

TAAACCACCA ATAATATTCC CTTCTCCTGA ATTATACTGT GTTGGAAGAG CAGATTGAGT 25920

CACACTCATG TAATAGCTAA AGGATAAAAA ATCTACGGTA TAATTTTTTA ATAACTCTGC 25980

ATCTTCAGCT GCAAACTCTA TGTTAATGTC ATTTTCCTTA AAATATCTTT TTGCATAATT 26040

CGGATAATAA CCTCTAACAT GCACATCTGA AAATAGATAA TTTAGATTCT CATACTCATG 26100

AGTCGCCCAT ACATCTTTTG GATTTGGAGT CATTGGATAA GCTGGCATAG CTAATACCAT 26160

ACATCCCACC TTAAACTCTG AATTAATCTC ACGAGCAATT TTTGTAACCA AACTTGAGGC 26220

GACTAATTCA TGATGTATAG CTTGATATAA TTCTTGTTTC GAAAGATTCT CCTTAGGTAT 26280

ATCTATTCCT CCACTAGTAA ATGGTAATTC CAAAACAGAG TTTACTTCGT TAAATGTAAG 26340

CCAATATTTA ACTTTATCTT TATACCTTTC TAAAACTGTT CGAGCAAATT TTTCATAAAA 26400

ATGAATCATT CTCCTATCAA CCCATCCATG ATATTTTCTT GCTAAATATA ATGGAGTCTC 26460

ATAGTGTGAA AGAGTTACAA GTGGTTCTAT CCCGTGAGCA TGTAGTTCAT CAAACAATTC 26520

ATCATAATAT TTCAACCCAG CTTCGTTAGG TTCTTCCTCA TCTCCTTTTG GAAAAATTCT 26580

ACTCCATGCA ATAGAAGTAC GAAAAACATT AAAGCCCATT TCAGAAAACA AGGATATATC 26640

TTCCTTATAT TTATGATAAA AATCAATACC TATCAATTTT AAGTTATCTT CTGTAGGATT 26700

AGACCTCACC GACGACAATT TTTGAGAGAC CAGCAGCTGG TGATTCGACA CCCTCTACCT 23220

CGATCCCTGT AGTTGACATT TTTTCAGCCA ACTCTTGTGA TGGCACATCA ATGTCCACCA 23280

ATTCTTTTAA CCATTTATAA GATACAAGCA TAATTTAGTT CTCCAGAATG ACAGTTGTCA 23340

CTCTAGTTCT TTTCCTTTCC TATCATTTCA ATAGAAGAAT CCTCTTCTTA CCTTAATTTC 23400

TTTCTCAGTA ACCAATCCGT ATCTACTTTT TGACCAACCA TAAAATGATG TTGGCTAAAT 23460

TTTTCAAAAC CATATCGGTT ATAAAACGCT TGAGCTTTTG TATTATGCTC CCAAACACCT 23520

AGCCAAGCCC AAGAAAAACT ATTTTTTGTA GCAAGTTCAA GTGCGAATTC AAACAGTTGC 23580

TTACCTAGTC CAAATCCTTG GAATTTTTGT AGCACATAGA GACGTTGAAT TTCAAAAGCG 23640

TCCTCTAATT CTCTCTCAGT TTGAGCACTT CCCCAGTTGA CTTTGAGAAA ACCAGCTATC 23700

TCCTCCTCAT GCATAATGAA ATAGGTTTCA GAGTCAGGAT TTCCCAACTC AGTTGACAAA 23760

GTTTTCAGAC TATAAGCCTC TTCAAAGTAT TCCTGTAACT GCTCTTCCGT ATTATCATAC 23820

GCAAAGGTTT CACGAAAGGT TTGTTTGGCA ATTTTAGCCA ACACCTCAAC ATCTGCCATT 23880

TCTACTTTTC TAATCATTAT TTAAACTGTT CTGAGAAGCG GACATCTCCT TGGTAGAATC 23940

CACGGATATC GTTGATTCCA TAACGGAGCA TAGCTACACG CTCTTGTCCA AGACCAAAGG 24000

CAAAGCCAGA GTATACAGTC GCATCGATAC CACTCATTTC AAGGACACGT GGGTGAACCA 24060

TACCGGCCCC CATAATTTCG ATCCAACCTG TTTTCTTACA TACATTACAG CCTTCTCCAC 24120

CACACTTGAA GCAAGAAACA TCCACCTCAA CAGATGGCTC TGTGAATGGG AAGTAAGATG 24180

GACGCAAACG AATTTGACGC TCTTCACCAA ACATTTTTTG GACAATCAAC TGAAGCGTTC 24240

CTTGAAGATC AGCCATAGAG ATATTTTTCC CAACTACCAA GCCTTCGATT TGGTGGAATT 24300

GGTGACTGTG GGTCGCATCG TCCGTATCGC GACGGAAGAC ACGCCCTGGC GAGATCATCT 24360

TCAAAGGACC TTTAGAAAAA TCATGGGCAT CCATAGCACG CGCCTGAACT GGAGACGTGT 24420

GGGTACGGAG CAAGATTTCT TCAGTGATAT AGAAAGTATC CTGCATATCA CGAGCTGGGT 24480

GGTCTTTTGG AAGGTTCATA CGTTCAAAGT TATAGTAGTC TTGCTCCACT TCAAAACCAT 24540

CCACGACTTG ATAACCCATA CCGATGAAGA TATCTTCGAT TTCTTCACTG GTTTGTGTCA 24600

AAACGTGACG GTGACCAGTC GCAACTGGAC GACCTGGAAG CGTCACATCT ATACTCTCGC 24660

TAGCCAGTTG AGCCGCGACT TTCTTTTCTT CCAAGAGCTT AGCTGTTTCT TCAAAAGCAG 24720

CAGTCAAGAC ATCACGAGCT TCATTGACGT GTTTCCCGAT GATTGGACGC ATCTCAGCAG 24780

AAACATCTTT CATCCCTTTG AGGATTTCAG TGAGCGAACC CTTTTTACCA AGGACAGAGA 24840

CACGCAAATC TTGCATCTCT TTTTCATTTC CAGCAGTAAT CTGCTTCAAG CTAGCCAGCG 24900

GAAGGCTAAC GATTTCAGAT GTTGCTGTAT AGGTTACTTG GAGTCCCAAA CGAGTAAATA 21420

GGGCTTCAAG GATTCCCTTA GCATAGAAGA AATCAACTGG AACTGCTGCT GTTTGGAAAT 21480

CTTTTTCAGC AACCAAGCCT GTCAAGGCAA AGGCAAAGCT GTTGATCTCA TTTGGAAGTT 21540

CTTCTTTTGG ATTACCTGTT TGTTCAAAGA CTTTTCCAAT CTCATAAAGG GCCAAGTTTT 21600

TATTCTTACG AGCCACGTTG TAGGCAACGG TATCAAGGAT CCCTGAAATC ATATTTTGAC 21660

GGAGGACTGA ACGATCCACA GTCATTGGCC ACATGAGTTC AGTAAGGTTA CTTGGTTGAG 21720

CTGTGAACTC AACTGCTTTT TCAGGAGTTG TCAGAGCATA GGTGATGATT TCTGTCAAAC 21780

CTGCTCCTTC AGCAATGGTA CGAACTTGAC GGCGGAGTTT TTGTATCACA GTCAATTCAC 21840

CAGCTGTACC ATCGTCTTTT GGAAGGCTGG TTGGCAAGCG GTCATATCCA TAGATACGAG 21900

CGATTTCTTC AAAGAGATCA GCTTCGATTG TGATATCCCA ACGACGACGT GGTACGCTGA 21960

CTGTAAAGCT GTCTGCATTT CCAGAAAGAC CAAAGCCAAG ACGACGGAAG ACGTCTTCTA 22020

CATCAGCATA AGACAGCTCA GTTCCGAGGA CACGGTTAAC ATCAGCAAGG GTTGAAGAAA 22080

CTTCCACATC AGAGGTATCA AGCTCACCCG CTGAAACGAT ACCCTTACGC ACCGTCGCGC 22140

CTGCAAGCTC TGCAATCATG CTAGCTGCCG CATCAAGGGC TTCATTAACT GTTGCCACAT 22200

TAATTCCTTT TTCAAAGCGA GAAGATGACT CAGAACGAAG GTTCAGGCGA CCACTTGTCT 22260

TACGGATAGA TTTGCCATTA AAAACAGCAG CTTCAAGGAT AACACGACTA GATTTTTCAG 22320

AAATTTCTGT AGCCTGACCA CCCATAACAC CGGCAAGGGC TACTGGTTTG TCAGCAACTG 22380

TAATCACGAG GTCTGTCTCA GCCAAGTCTC GTTCTTCACC GTCCAGGGTC ACTAATTTTT 22440

CACCATCACG CGCTTCACGC ACACGGATGT CAGTCCCTTC AAATGTGTCC AAGTCAAAAG 22500

CATGCATAGG TTGACCAAAG TAGAGCAGGA TGTAGTTTGT CACGTCTACA ACGTTATTGA 22560

TGGGACGGAT GCCTTCGTTC ATGAGAAGGT TTTGCAACCA TTGTGGACTT GGTGCGATAG 22620

TCACATTGTC CAAGATACGA GCTGCATAGT AAGGCGCCTT GTCTGTCTCA ATGCTGACAG 22680

AAAGGGCATC TGCCGCAGCT TCATTAGTTT CTGTTAGAGT AAATTTTTTA AAGTTGACTG 22740

CCTTGTCATA GATGGCTGCC ACTTCGTGAG CCACTCCACA CATAGAAAGG GCATCTGCAC 22800

GGTTTGGTGT GATGGAAAGT TCGATGATTT CATCATCCAA GTCTAGGTAA GAAAAGACTT 22860

CCTCACCTGG CACGGCATCT TCAGGCAAGA TTTGGATGCC ATCTGCGAAT TCCTTAGGCA 22920

CAACTGAGTC AGAAATTCCC AATTCACCAA GTGAACAGAT CATTCCAAGT GACTCCAAAC 22980

CACGGATTTT TCCTTTTTTG ATTTTGTAGT TATCAGCGAT ACGAGCTCCT GGAAGAGCCA 23040

CCATGACCTT GATCCCAGCA CGCACATTTG GGGCACCACA AACGATCTGA CGCTCTTCTT 23100

CTTCGCCAAC GTTAATCTGA CAAACATGGA GGTGAGTCTC TGGCACATCT TCGCAAGACA 23160

CGATTAAGCG ACGGCTATCG ATTGTCAAAT GTTTGCCTTT TGTAGTATAA TTGTCTTGCA 19680

TTTCTGTGCC TTTTAATCAT TTCAATCTTA AATTGGACTT TTTTTACTTG GGTTGTACTT 19740

AATCTATGAG GAAGACAAGA AAAAGAATAT CAATCAAGTA AAGTCACAAA GTCACATTAG 19800

CTCCGAGCAA CCATTGCAAA TTGAGGTACT CACACAATGA TTAAAACATT TCTCTCTGCC 19860

CTTTCGGTCA TTCTCTTTTC TATCCCTATC ATAACTTATT CTTTTTTCCC ATCTTCTAAT 19920

CTTAACATTT GGCTATCTAC CCAACCTATC TTGGCACAGA TTTATGCCTT CCCCTTAGCT 19980

ACTGCAACTA TGGCTGCTAT TTTAAGTTTC TTATTTTTTT TCCTATCTTT TTACAAGAAA 20040

AATAAACAAA TACGGTTTTA CTCTGGCATT TTGCTCTTAC TATCGCTCAT ATTACTATTA 20100

TTCGGAACAG ATAAAACCCT TTCTTCTGCA TCAAATAAGA CTAAAACCTT AAAATTAGTA 20160

ACTTGGAACG TCGCTAATCA AATAGAAGCA CAACATATTG AGCGAATTTT TAGCCATTTT 20220

GACGCCGATA TGGCTATATT CCCTGAACTA GCTACCAATA TCAGAGGTGA GCAAGAAAAC 20280

CAGAGAATCA AACTATTGTT TCATCAAGTT GGACTTTCTA TGGCCAACTA TGATATTTTC 20340

ACTTCTCCAC CTACCAATAG TGGAATAGCT CCTGTGACTG TGATTGTCAA GAAAAGTTAT 20400

GGTTTCTATA CAGAAGCTAA AACTTTTCAT ACAACACGGT TCGGGACAAT TGTATTACAT 20460

TCGAGAAAAC AAAATATACC AGATATCATT GCCTTGCATA CTGCGCCTCC TCTGCCAGGT 20520

TTAATGGAAA TCTGGAAGCA AGACTTAAAC ATCATTCATA ATCAATTGGC TTCAAAATAT 20580

CCAAAGGCTA TTATTGCAGG TGATTTTAAT GCAACTATGC GTCATGGAGC ACTTGCAAAA 20640

ATAAGCTCTC ATAGGGACGC ATTAAATGCA CTGCCACCTT TTGAAAGAGG AACTTGGAAT 20700

AGCCAAAGTC CAAAACTTTT TAATGCAACA ATAGATCATA TTTTATTGCC TAAAAACCAC 20760

TACTATGTTA AAGATTTAGA CATTGTAAGT TTTCAAAACT CTGATCATAG ATGTATTTTT 20820

ACAGAAATCA CATTTTAATT ATTTTATATA AAATCACCCC TCTAATGTTC ATAAACTAGA 20880

GGGGGAATTT GTATCCTACT ATCGTTTAAC GCACTTCTGC ATTGACTTTT TCTTCGAGAG 20940

ACGCTTGGAT TTTTTCCATA TAGCGTGCGA CTTCTTCGTC CGTTAAGCTG TCTTCTGGAT 21000

TTTGGAAGGT CAAGCTATAA GCCATTGACT TCATACCAAG TCCCAGTTTT TCACCTGAGA 21060

AGACGTCAAA GAGTTTGATA TCTGTCAAAC GTTTCACGCC GGCAGCTTGG ATAGCATCTA 21120

CAACTTCTTG GTGAGTCACT TCTGCCTTGA GGAGAAGGGC AACGTCACGG CTGACTGCTG 21180

GGAATTTGGT GATTTCCACA AATGGAACAG CAGGTTGGAG CGCCCCTTCG ATGGCTGAAA 21240

GGTTAAGCTC AGCTACATAC GTTTCTGGAA TATCGTAAGC CTTGGCAGTG ACTGGATGCA 21300

CTTGGCCAAG GAAACCAAGA ACTTGGTCAC CGAGTGAAAT CACGGCTGTA CGACCTGGAT 21360

AGTCAAGACC GATAGCATCT ACTGGTAAGT TTACAAGGTC AGCGTATACG TCACGAACAT 17880

CACCGAAATA AGTTTGAAGC AAGACTTCAA GACCTTTTTT GTCAGCCAAG AGTTTGTTGT 17940

AAAGGTTCAA GAAGAGAGCT TTTTCTTCAG CTGTCAAGTC TTTTACAAGA GCCGCTTCAT 18000

CCAATTGGAT GCGAGTCGCA CCAAGTTCAG CCAATTTAGC AAAAACTTCT TGGTAAGCAG 18060

CCACTAAGCT ATCTACGAAG TCGTCTGCTT TCACGCCTTC TTCAAAGTCT GACAATTGAA 18120

GGAAAGTGAA GGGACCTACA AGAACAGGAC GAGTGTTCAA TCCAAGTTCT TTGGCTTCTT 18180

GGAACTCATC GAAAATCTTG TGACCAGCCA ATTTTACTTG AGTGTCTTTT TCAAATTTAG 18240

GAACGATGTA GTGGTAGTTA GTGTTGAACC ATTTCTTCAT TGGAAGGGCG CGAACGTCCC 18300

CTTTTTCTCC CTGGTAACCA CGTCCCAAAG CGAAGTAGCG CTCAAGGTCA GACAAGTCCA 18360

AGTTTTGAAC GGATGCAGGC ACCACGTTGA AAAGGAAAGC CGCATCTAGG AAGTTATCAT 18420

AGTGAGAAAA GTCATTTGAT GGAATTTCAG TGATGCCTTT TTCTTTGACA ATGTTCCAGT 18480

GTTTAGCACG CAAGTCTTTT GCTGCTGCTA AAAGTTCTTC TTCTGAGATT TCTTTTCTAA 18540

AGTATTTTTC AGTTGTAAAT TTTAATTCAC GGAATTCGCC CAAACGAGGG AAACCGATGA 18600

TTGTAGTTGA CATGATGTGT CCTCCAAAAT TTGTTGTTGA AACTATCTTA ACAGAAAAGA 18660

AAGCGTCTGT ATAATTGTAA AAAATTAGGG TTTGATATAG TTTGAAACTA TATATCTGTT 18720

TCGGACAAAA GAAAAAGACT TGAAGCAAAC GTCTCAAATC CTTTGTAATT CTTACTTTAC 18780

AGCTATATTC CAATTAGAAT ACTAAAACAT GTTATTAGTA ATTCTTATAA GTGACTATGA 18840

CCTGTTATTA GAAAAGACTA TAACTGATTC TAGTCAACTT TTTCCCTGTT CAAGTGGGAC 18900

GATTGCTAGT GTCTTTCCTA AACTGGCTAG GACTTTTAAG ACTGTATCCA ACTGAGGACT 18960

AGTCTTTCCT GTCTCCATCC TAGCTATGAC AGGCTGGCTT ATTCCACTGA CTTCTTCCAG 19020

CTTTTTCTGA CTGATTCCTT GTTCATACCT AGCCTCAATC AACTCGCTCA TGATAGCCAC 19080

TCGCATATCA CTTTCAAGGA TTTCCTCCTT GCTAAAGAGC TCAGATGGAC ATCCTTCCAA 19140

TTACTCCCAA TAGCACTATT CTTCATCACT TAACCCTCTT TTTTTTACGT CTATGTATTT 19200

TTAAAAAAAT GAGCGAATTA TGATTCGATA GATTGACCAG TGGGTTTAAA GTTGGTGCTA 19260

GCCTATTTCT TAAGCGATTT TCCTTTTCTA GGATAAAGCA GTTCCTGCTT GCTTAACCCC 19320

AATTTTCCAC GATGAATCCA ATAGTAAATG GTTGAAATTC CCACGTTAAC CCCTTTAGCC 19380

ATCACCATCA TTTCAGGCGA AAATTTTTGG TTATGTTTTT GGTTATGTAT AGTGGAGAAT 19440

CTTTTCCTTT AGTTTCTTAA GACTGTTGAG CGTAGTCGGC AGAATAAATC TCTTTGAAGC 19500

GCCCTTTTCC AAGACATTGT CGGACTGTCC CACGCTTGAT TTCAGTGTGG ATAGTTTGAG 19560

GAGCTTTTCC AAGTAGAGAG GCAATTTCTC TATTTGATTT TCCTTCTTTT TTCCATCGTT 19620

GCAAGTGGGC AATGGTCGGA ATCGCCAAAT CATTTTGGAT AAAGTCAGCC AAACGAACCG 16140

TCGTTTCCTT GATATTAAAT TTATTATTGC TGGCAGTTAC ACTGATAAAA TGGGGAGCCA 16200

ACTCCTGCAT ATCCTGCAAG GCTGAAATAA TGTTATCATT ACCCACGGCT GGGTTTGGAG 16260

GGAACACTTC AAATGAGAGT GACGGTGTTT GGCGTGACAT ATGTAATAAC CTTTTCTAGT 16320

TGATTTCTTT TTGAACAACC ACTGTATGGA GAGAAATCCA ATCTTACAAT TTCTCACGCG 16380

CAGCTTTAGC TGCTTCAACA AGGCGGATCA AGCTTTCTTT TGTTTCTGGG ATACCACGTG 16440

TTTTCAAACC ACAGTCAGGG TTGATCCAAA CTTTCTTGCT TGGCACTTTA GCAAGGATGG 16500

CTTCGATTGT GTTGTCGATT TCGCCTTCAT TTGGTACACG AGGTGAGTGG ATATCGTAAA 16560

CCCCAGGTCC CACTTCTGTT TGGAAGTTTT TCGCTTTGAG TTCGTCCAAG ATTTCAAGGT 16620

TTGAACGGTT AGCTTCAAAG GAAATAACGT CTGCATCCAT GTTATCGATA GCTGGGATGA 16680

TATCTGTAAA TTCTGAGTAA CACATGTGAG TGTGGATTTG TGTGTCTGGC GCTACTGTTG 16740

AGTGTACCAA GCGGAAGGCA GGAATAGCCC AGTCAAGGTA GTCTTCGTAC CAGTCGCTAC 16800

GGCGGAGTGG CAATTTTTCA CGAAGAGCAG CCTCGTCGAT TTGGATGATT TTCACACCAG 16860

CAGCTTCAAG GTCAAGTACT TCATCCTTGA TAGCAAGGGC GATTTGGAGA GTTGAATCCT 16920

TGATAGAGAT GTCTTCACGT GGGAATGACC AGTTAAGGAT GGTAACAGGT CCAGTCAACA 16980

TACCTTTAAC AGGTTTGTTT GTACGACTTT GTGCATAGCT AGACCATTTA ACAGTGATAG 17040

GGTTAAGACG AGTGACATCA CCCCAGATGA TTGGTGGTTT TACCCCACGC ATACCGTATG 17100

ATTGTACCCA TCCATTTTTA GAGAAGAGGT ATCCTGACAA GTTTTGACCG AAGTACTCAA 17160

CCATGTCATT ACGCTCAAAT TCACCGTGAA CAAGGACATC AAAGTCAATA TCTTCTTGCC 17220

ACTTGATCCA TTCGTCAATC GTTTCAGCAA GGAAAGCGTC GTACTCTTTT TGAGACAATT 17280

CACCTTTACG GTAAGCCAAA CGTTTGGCAC GAACTTCTTT TGTTTGAGGG AATGAACCAA 17340

TCGTTGTTGT TGGAAGAGCT GGAAGTTTGA AAGCTTCTTC TTGGATAGCT TCACGTTCTG 17400

CAAAGGCTGG CAAACGAGTG TAGTCTGCGT CTGTCAAGCC AGCGATACGC GCACGAAGTT 17460

CAGCATTTTC ACCAACACGC TCAGTCGCAA AGAGTTCTTT GTTGGCTGCA AGAGCTTCTG 17520

AACCTTGACC ATTTCGGATA GCATCCAAGT CACGGATTTC ATCCAATTTT TCAACTGCAA 17580

AGGCAAAGTG GTTCAAGAGT GCTGGTTCAA ATTCTTCATT AGCAGTTGTA AATGGCACAT 17640

GAAGAAGTGA GCAAGAGCTT GTCAAAACGA TGTTTTCAGC TGGAATTTGC TCAAGAACAG 17700

CCAAGCTCTT TTCGTAGTTG TTGCGCCAGA TGTTTTTACC ATTGACAATA CCTACATAGA 17760

GAGTCTTGTC AGCTGGGAAG CCACCTTTAA CGAGTTCAAG AGTTTTCTTA CCTTCAACAA 17820

CATCTACATA TCCCACTTGT ACCCCAGCAA TTGGTCCGTC AAATGGAATA TCTGAAATAG 14340

ACAGTGCCAA AGATGAACCA AACATAGCAG CCATTGGTGC AGATGCATTT TCATCATAAG 14400

AAAGCACTGT ATTGATGACT TGGACTTCAT TACGGAAACC TTCCGCAAAC ATAGGACGAA 14460

TCGGACGGTC AATCAAACGC GCTGTCAAGG TCGCATCTGT TGAAGGACGT CCTTCACGTT 14520

TCATAAAGCC ACCAGGAAAC TTCCCAGCCG CATACATTTT TTCTTCGTAG TTGACTTGGA 14580

GTGGGAAGAA ATCCCCAGTT GCCATTTTCT TAGACATAAC GGCAGCAGTC AAGACAGTTG 14640

ACTCACCGTA ACGTACGACA ACAGATCCAT TTGCTTGCTT AGCAACCTGA CCAGTCTCTA 14700

CAATTAACTC ACGACCCGCA AAAGTCGTTT GAAACACTTG TTTTGCCATT TTAATCCCCT 14760

TTGGATTGAT GAAATTATAC GCCTTGCCTA CAAAGATCAA GATACCAAGG ACGTCAAAAG 14820

CAAAGTAAAA ATAGGAAACT GACGAAGTCT TCGATGAAGA CAAGACAGTT TATCTTTTTT 14880

ACACAGCTTT TCGGCCGTGT TCAATTACAC AAGATATTTT GGACGGTTCG GCTTGCCGAA 14940

CATTTCTGTA GAAAAATAGG AAGGTGACGT CGCACTCGAC GAGTGCTAGG AAGCTTATCT 15000

TTTTTCCTAA GAAATGAGAC CAAAATTCAA GTCATCAAGA TACCAAGCCG TCAAGCAACT 15060

CAAAGGAAGA TAGGAAATCG AACGACGGAG CGACTACTCC TAGGGAGATT TATCTTTTTC 15120

CACAGAGTTG TAGGCAAGTT CAGTTTTCAA GATACATCAT TAGAAAGGTT TAATACTAAA 15180

GTATCTAAAG CTTTCACGCT AATCGCTATC GGGCGATTAG CTAAATGCTT TACTAACTCT 15240

CTCGTCAAAT AACATCGATT TGACTCACTC GTGTCGTTAA ATCTTACAGT TTAAATGCAT 15300

TGTATTATTT AATACCTTCA TCTTTGTATC AAGTACGTAC AGAATTTATT TTATCATATT 15360

TTTCTTAAAA AGTGAGGTCT TTACCATTAA AAAGGAACCA TTCCCCTCAC CTGAGAAGAA 15420

TGGTTTGCTT TTATTATCCT AGAGACTGGT GATTAAACAA GGCATGGGTT GCTTGATGGA 15480

TGTATTTTGC TGTATCAGCA TTATTCATCG TATAGAGATG CACACCGGCA ACATCCTGAG 15540

TTACCAAGTC CACGATTTGG TCCACTGCAT AGGCAAGTCC TGCTGCTCTG AGCGACTCAG 15600

GGTCATGCTC ATACTTGTCT AAGATGGCTT TAAATTTGCG TGGAAGATGG ATATTCTCAC 15660

AAGTCTTCAA GAGTCGGAGA GCCTGATTTC GATTCAGAAT TGGCATAATT CCTGCATGAA 15720

TGGGAACATC AATCCCAGCC AAGATACACT TGTCCTGAAA ATCATAGAAG CGCTCATTGT 15780

CAAAGAAGAG CTGAGTTACG AGGCTCGAAC AGCCTGCATC CACTTTCTTC TTAAGATTTT 15840

GAATATCTGA AATCTGATTT GGCGAATCTG GATGCCCTTC TGGATAGCAA GCTCCAATAA 15900

TATCAAAGTG AGGGGTTTGT TCCTTGATAA ACTCAATCAA GTCGGTTGCA TAGCGGAAAT 15960

CCTTTTGTGG TTCCACGTCT GGAATAATAT CCCCACGAAG AGCCAAGATT TTCTGCACCC 16020

CAACTTTGTC CAAGTCAGCA ATAGTTTCAG CAACCTTGTC CTTAGTTAGA TAAATAGCTG 16080

CTGAATCTTT TGATGTTTCT GTAAATTCTT TCTTAGGTTT GTAATCCTTT TGATGACGTG 12600

GGCGGTGAGG GCGCTCAGAC TTTTCACCTT TTTCATCATG CTCAGGTTTT GGCGGACGAG 12660

GTAGAAGAGC CTTCATAGAG GCATCGATAC GGCCTTTTTC ATCAATTTTG ATAACCTTAA 12720

CATCAACTTC ATCCCCGATT TCTACCAAAT CCTCTACACG ATTGGTACGA GTCCAAGCCA 12780

TCTCAGAGAT ATGAACAAGG GCATCTGTCT TATCAAAGAG GTTAACAAAG GCACCAAATT 12840

TCTCGATACG AACGACTTTA GCACGGTAAA CTTCATCCAC TTTGGCTTCA CGAACCAAAC 12900

CAGCAATAAT TTCTTTGGCA CGGTTAATAG CATCTTGGTC ACTAGAGTAG ATAGACACAT 12960

TTCCTTCTTC GTCTATATCA ATCTTAACAC CTGTTTCAGC GATAATCTTG TCGATGGTTT 13020

CTCCACCCTT ACCGATGACA ATCTTAATCT TGTCCACATC AATCTTGATC GTATCAATTT 13080

TCGGAGCAGT TGGAGCCAAT TCTGGACGAA CTTCTGGAAT GGTTGCTTCA ATGACATCAA 13140

GGATTTCAAA ACGCGCTTTC TTGGCTTGAG CAAGAGCCTC CGTCAAGATT TCTGCAGTAA 13200

TCCCTTGAAT CTTGATATCC ATTTGAAGGG CTGTAATCCC ATCACGAGTA CCTGCAACCT 13260

TGAAGTCCAT ATCTCCAAAG TGATCTTCCA AACCTTGGAT ATCTGTCAAT ACTGTGTAGT 13320

TATTTCCATC TGAGATAAGC CCCATAGCAA TACCAGCTAC TGGCGCCTTG ATTGGCACAC 13380

CACCAGCCAT AAGGGCAAGA GTTCCCGCAC AGATAGAAGC TTGAGATGAA GAACCGTTTG 13440

ATTCCAAAAC TTCTGCTACT AGACGGATAG CGTAGGGGAA TTCTTCCAAG CTTGGCAAGA 13500

CTTGAGCAAG AGCACGCTCA CCAAGGGCAC CGTGACCGAT TTCACGACGA CCTGGCGCAC 13560

CGTAACGACC TGTTTCCCCT ACAGAATATT GAGGGAAGTT ATAGTGGTGC ATAAAGCGTT 13620

TCTTGTACTC TGGATCCAAA CCATCAATGA TTTGAGTTTC TCCCATCGGA GCCAAGGTCA 13680

AGACTGAAAG AGCTTGAGTT TGCCCACGAG TAAAGAGACC TGAACCATGT ACACGAGGAA 13740

GGAAGTCAAC AACCGCATCC AAAGGACGGA TTTCATCGAC CTTACGACCA TCAGGACGCA 13800

CCTTGTCTTC TGTAATTAAA CGTCGCACTT CTGCGTGTTC CATTTGTTCC AAGATTTCAG 13860

CCACATCACG CATAATACGG TCAAATTCTT CGTGGTCCGC ATATTTTTCT TCGTAAACGG 13920

CAGTCACTTG GTCTTTCACT ACTTGAGTCG CAGCTTCACG GGCCAATTTC TCTTCTACTT 13980

GAACTGCCTT TTGGAGGTCA CTGTTGTAGG CTGCAATGAT TTCAGCTTGC AATTCAGCAT 14040

CCACGTGAAG CAATTCCACT TCTGCTTTTT CTTTACCGAC AGCAGCAACG ATTTCTTCTT 14100

GGAAGGCAAT CAATTCTTTG ACAGCTTCGT GCCCTTTAAG GAGCGCTTCC AACATGATTT 14160

CTTCTGACAA TTCTTTGGCA CCAGACTCTA CCATGTTGAT AGCGTGCTTG GTTCCAGCTA 14220

CTGTCAATTC AAGAAGAGAT TGCTCTGCTT GTTCTTGACT TGGGTTGATG ATGATTTGGC 14280

GCAACCTTAT ACCCACGGTA CTCAAAATAG CGACGAATCG TATCAAAAGC TACCGTCGAA 10800

CGGGCGTTTC CTACGTGGAT ATAGTTGTAC ACCGTTGGCC CACAAACATA CATCTTGATC 10860

TTGCCGTCCT CAATCGGGAC AAATTCTCGC AAATCACGAG ACATGGTGTC ATAGATTTTA 10920

ATCATAAATC ATAATCAGGA AAGCTGAAAT CCAAGAACAA TTAGTTTCAT CACTAAAAGT 10980

TCAAGTAAAT TTCAGTCCGA ATATCTCTAC ACTTCGGAAT CeeTTGCTCC TTTCTCATTC 11040

AGATAAACCA CCTGAGTCTG TTTGACAAAG CCAATTTTTT CATACAAACG TTTGGCACCT 11100

ACATTGCTAT CTTCCACTGC AATCTGAAAT TCCTTGTCAT TTTGCTCAAT TAGTTGGTTG 11160

ACGAGGGATT TTGCTAAGTA GCTTCCATAG CCTTTTCCAC GTTCAGGTTC CAATATTGCT 11220

AAACCGTAGA GGTAATTCGT ATTAGTCGAT AAATCAACCG TACAAGTTCC AATAACCTGA 11280

CCAGCTTTTA ATAAAATATA TAGTCGGCTT TCTGGATCTT TCAGAGCTTC AGCGACATAT 11340

CTATCCACAA CTTCTCTCGA TTCATGTTCC TCTGAAAATG CCTGAAATTT TAATTGACTA 11400

ATTTGATCCT GATACGAACT ATCTGCTAAC AAAACTTCAA GATGGGAAAC ATTTGCTAAC 11460

GGATAAGGTC TTCTATCCTT ACCTAACCAA GTTTCTGTCT CTTCATCCTC GATTAGTCCC 11520

CAGTTACTGG CAAAGTCAGG ATGATTCTCT AAAAAAATAC GTTCTGTCTG AAAAGTGACT 11580

GACCGAATGG GGAAAGAAGC TGTTTCTCTC TCAAAACTAG TAAACAATGC ACGCGCAATC 11640

CCCTGACGGC GATGACCTGG ATGAACCAGT ATCGTCACTT CTACATCTTG GTCATCTGCA 11700

TAGACAGTTA ATAAACCAAC AAGTTCGCCT TTTTCATAAT AAAGGAAAAA GGCGGGCATG 11760

TTTGGGTCAA AATTAAGCAT GTTAGAGAGA TAGGGATCGC GATAGGTACC GTCATAGTTT 11820

TGGCAACAGT TAATTACTTT TTTCGCCTCA GATAGCTCCT CTTGGCTTAA CTTGTTTCTT 11880

GCTTGAATCA TATAGGTATC CTCTACAAAC CAGACGATCT GTGACTGGCA TCTTTAGCCT 11940

GCTCGAGTTT ATTGACATAA TACTCTCGTT TTTCTTCGAC TTCGTGAATG ACAGGCTCAT 12000

CTTTCTTACC ATGAAGACGG ACAATCTTGG CCGGAATACC GACAACCGTC ACGTCACTAG 12060

GTACATCTGC TACGACAACT GCTGCAGCAC CGACCTTGGC ATTTTCACCA ATTTCCACAG 12120

GCCCGATAAC TTGGGCATGG GCTGATATGA GGGCTCCCTT TCGTACAGTC GGATGGCGTT 12180

TGCCACAGTC TTTCCCTGTT CCCCCGAGAG TCACTCCGTG ATAGAGAAGA ACGCCTTTTT 12240

CAACAATCGC TGTCTCTCCA ATCACCAGAC CAGAACCATG GTCAATAAAA ACACCTGAAT 12300

CAATCTGGGC TCCTGGATGA ATCTCAATCT GAGTCCAAAA GCGCCAAAAC TGACTGTACA 12360

TACGAGCTAA TAGTTTGAAG CCGTGCTTCC AGAGAAAATG CGAGAGACGG TGGGCCGCCA 12420

AGGCCTTGAC ACCTGGATAA GTCAGCAAAA CCTCCAAAGT GGTGCGGGCC GCTGGATCAT 12480

TTTCTTTTAC AATATCAATG GTTTCGCGCC ACCACCCCAT ACATTTCTCC TTTTCTTATT 12540

ACCAAGGGTT TAGCGAACGA GCCTGCTCCT GCTGGGTCAA AAGGGCAACC AACTGCTTTT 9060

CACGCTCGCT GAGCCCAGCT TCCTCCAGCA AAATCCACTG CTGAGAAGCT AAAGGGAGCG 9120

TGAGATAGCC CTCTTTCTCT ACTGGTTGGT CTGAAATCCG AGCCTCAGGA AACCAGTCTT 9180

GTAGTTCTTT TGCCCTCATG TTCTAGCCCT CCACTTTTTG GATGCACCAT GAAACCAAAC 9240

TCTCAAGACG TTCCAGATTC TCAGTCATAT GGAGATAGCC CATAACCGCT TCAAATCCCG 9300

TGGACATACG ATAAGTCACG ACATCTGCAT TTTTAGCCTT TGTGTGGCTA TTGGTATTGC 9360

GGCCACGTTT GTAGATTTCT TCTTCTTTTT CCGTTAGGAC CTGCTCCTCC AACATGAGAG 9420

CAATCAGGCG AGCCTGAGCC TTGGCTGACA CGTACTTAGT TGCTTCTTGA TGGAGTTTAT 9480

TGGGTTTGGT CATACCTTTG AGGATGAGGT GACGGCGAAT ATACATAGAA TACACCGCAT 9540

CCCCCTCAAA GGCTAGCGCA ATCCCGTTAA TGAGATTGAC ATCAATCACG TGTCCACCTC 9600

ACTCCATCCT TGGTATCAAG GAGCTTAATT CCTTGAGTAA CCAATTGGTC ACGGATTTGG 9660

TCTGCTGTCG CAAAGTCACG ATTGGCACGC GCCTCTTGGC GTTTTTGAAT CAAGTCTTCA 9720

ATCTCTGCAT CCAAAACTTC CTCAACAAAG ACAATTCCAA AAATTTCTAA CATATCTGCA 9780

AGAGCTTGCT TGACACTTGC ATCATAGTTC CCTGAGTTGA TCCATTTGGC CATTTCAAAG 9840

ACAACTGTGA TACCGTTGGC AGCATTAAAA TCTTCATCCA TAGCTGCTAC AAACTTATCT 9900

TTAAAGTTTT GTAACTCTTG GGCATCCACA TTTCCTGTAA ATGGTTGTTC GTAAGTATTC 9960

TTGAGATACT TGAGATTGGT CTCGGCATCG CGAACTGCCT TTTCCGTGAA GTTGATAGGC 10020

TTACGGTAGT GCTGGGTCGC AAAGAAGAAA CGAAGTACTT GCCCATCAAG AGTTTTAAGG 10080

GCATCGTGTA CCGTAATGAA GTTACCCAAG GACTTAGACA TTTTGACATT GTCGATATTG 10140

ACAAAGCCAT TGTGCATCCA GTAGTTAGCA AAAGCCTTGC CTGTTTTAGC TTCAGACTGG 10200

GCAATTTCAT TGGTGTGGTG TGGAAACTCT AGGTCAGCTC CACCACCGTG GATATCAATG 10260

GTATCACCTA AAATCTCTGT CGACATGACT GAACACTCAA TATGCCAACC CGGACGTCCA 10320

GGTCCCCAAG GACTATCCCA AGAAATCTCA CCTGGTTTGG AAGATTTCCA TAGAGCAAAG 10380

TCTACAGGAT TTTCCTTACG AGCCGTTTCT TCATCGGTAC GACCTGAAGC ACCTAGCTCC 10440

AAATCTTCCA AGGTTTTATT AGCCAATTTA GCATAGTTGT GGGATTTTTC TACACGGAAA 10500

TAGACATCCC CTTGACTCTC ATAGGCAAAG CCTTTCTCGA TCAAGTCTTC CACAAAACGG 10560

ATGATGTCTG CCATAAACTC CACTACACGC GGATGGCGAG TCGCAGGTTT CACGCCCAAT 10620

GCCGTCACAT CCTCACGAAA GGCAGCGATG TACTTATCCG CAACCTCCTG AGGCGTGATA 10680

CCTTCTTCCC TGGCACGGTT GATAATCTTA TCATCCACAT CTGTAAAATT GGAAATATAG 10740

AAAATATTGT TAGAAATCGA TTTGACTGTC CTGAACAATT CGTCCTATTC TTATTTCATT 7260

TTACTATAAT TGATAGTGGT CGCCCCAGCC AGATACCTTA TCTGCTATCC ATTTAGGAAC 7320

CCCTAACTTA AGCAATCCCC ATAATCGTCT CGATTTCTTC TTCCATTGCT TCCAGATAAT 7380

CACTCGTAGG CGAGTACGCA AGCGCTCATC TATGCTAGTG ACTATACTTT TCATATTTAT 7440

AATTCATTCC TTTCGTTTCA CTCAAGGCAC AACACAGAAT GAAAAAGTGT TGTGATCTTT 7500

ATTTTGTTTT ATAATAATAG TGAGAAAACC TATCACTACT ACAAATCACG GGGAGGTGAA 7560

TAAGTGAGTG GTACAGCCAC TACCTCGCAT ATTTTGTCAC ATCATTTAAC GGTACATAAT 7620

AAGTTGTACC ATCTGAATAA GTTGCTACAA TATCATTTGC ATGCTCTCCT TCACCTTTAG 7680

CAAAGGTTGG AGCTCCTGCT GGATGATTTT TATTTGCCTC TTTCAATTTT TCAATAATGG 7740

CATTTTTTCT GTATCTTTTA TATTATCAGG ATTTTTCACT AAGATTTTGT CTGGATATGT 7800

CGGTTTAGCA GAAACAATTT TTACTGTTAC TTCTTTTTTA TTCGAAGCAC TTGTCCAGTT 7860

TCCAGCATTA TCTTTAGCAT TTAATTTTAC AGTAATTCCT GAACTAGGAA CTTCAGTAGC 7920

AGGTTGATTA TCAACATTAT TCAACTTTAA TTTCAAAAGA GCTGTTGCAT CAGACGTTTT 7980

ATCAATCGTT ATATATAATG ATGAATTGTT ATTATAAACA GTTCCTTCAT ATTTAGCTGT 8040

TTGTGAGCTA CTTGAAACAG AACTGAAATT ATACCCACTA CCTCCCTGAT TATCTTCAAT 8100

GCTTACGTCT AAATGAACTT CCCCACTATT ATTTGGCTTA GCAACAACTG TTATAGTAAA 8160

ATAACATAAA ATTTGCATAA ATAGATTAGG GAAATCAAAG CAGCTTCTAG GAATGTTTTA 8220

GCAGTCACAG TGTACTTTCC CAGCATCAAG CCACTATAAC TCTGCACATA AAAATGGAGA 8280

AGATGGCAAT CCTCTTCTCC AAATATTAAC TTCTTTACAA ACCAACTATA GTTGACAAAG 8340

AACCTAAAAT CAATTGATAA CACAAGGTCA GGTCGGTCAA CTCTTTCAAC TGAAGCCCTG 8400

TCAACTCTTC CCATTTATCA ATCTTGTATT GGAGAGAATT GCGGTGCAGA TAGAGTTGCT 8460

GGGCTGTTTT AGTGAGAACA GCACTATTTT CCCAAAGAGA GAGAATGATT TCCTGAATCT 8520

GATCTTGATC CAAAATCATC TGGTGTAGAC ATTCCTTGAT TGGCTTCAAG TCCACGAGTC 8580

TTTCTCCCAT ACTCCAAAGA TAGAGCTGAG AAAAAGTATG AACACCTTGG TGACCCTGAC 8640

GCCACCATGT CTTGAACAAA TCCCGCTCAG CTTTGATTAA GTCTGATAGG GCTTGATGTC 8700

CCGTCTGAGA CCAAACCTGA CCCAACATGA TAGAAAGACG AAGTCCAAAG TCATACTCAA 8760

CCGCTTCAAT CGTATCACTT AAAATATCTC TTACAGAAGT GTATTTGTCT TGTTGAAGCA 8820

CGAAAACATA ATCCTGAGCT CCGACCTGTA GCACTGTCTG ACAATTCGGA AAAAGAGTCC 8880

GCATCATATC TAGCCAAGAA GCCAGATTTT CCTGCTGAAA ATAAGAAAGA TGGCAATAAA 8940

CCAACTGAAT CTTTTTAAAA ACTTGCGGTG CCTGTCCCTT GCCCTCAACC AGATAGGAAT 9000

GACCAGCAAG TAGGGACAAG AGTGTGGATT TTCCTGCTCC TGACTCCCCA ATAATACTGT 5520

AAAATTTTCC GGGTTCAAAA TTATAATTGA TCTGATATAG GACTGCTTCA GCAGTATTCT 5580

TATAACGGTA GGTAACATCT TGTAATTGTA ATAAAGTCAT GATTTCTCCT TCTTAACTAA 5640

TAGATGATAA AATTTCTTTC GGTGATTTTC TAAATAAGAA TAGGAAACAA AGGGCTACAG 5700

ATAAGCAACT AAGCAGAACT AGAAAAACAT AGGATTCTGC AAAAGATAAG ATGCTAGTTG 5760

ATAAACTGCT TGCTTTGGCT AGTGTATCTT GTAAGCTTGC CTGATCTCCA CTTGCTAGTA 5820

GAGTTTGGAG TAGGTAAGTT GTGATTGCGT TTCCTGCAAC AAATGCTGGA AGCAAAGCTC 5880

CAAGAGATAC CAAAACTACC TCTAAACAGA ATTGTAGGAA GATCGAGCTC TTGCCTTTTC 5940

CAAGTGCAAG TAAAATCCCC ACTTCATAGA CCCGTTCTCT CAACCAGAGA GACAAAACCA 6000

GAATTAAGGC TCCAGCTCCT GCTATCAACA TCCCATAAAG GAAGATGGTC AGGAAGGTTT 6060

GGAAAGTTGC AACTGAGTCT TTGATTTGTT CAAAAGCCTT GTTTTCCTTT TCGACTTGGT 6120

AGCCTTGATT TTCCAAGGCC AAGTTTTCTA CCTGCTTCAT GAGTCCGTCC ATTTCCTTAG 6180

GATTTTCTAC ATAGAAGCGT GCTGCACTGA CTTGAGCTTC ACTATTGCCC AAAAGGGTTT 6240

GGCTACTTTC ATAGTCTGTA AAGACTTGAT TTTCACTGAA GTCAGAAGAC AAGCCTGTGA 6300

ATTTCTCTTG TTTTTTACCA GAAAAGATGC CGATAATCTC AAACTCTACT GTTTGTCCTT 6360

TTCCAGATTC AGACTGACCA GCATCCAAGC CAATCTTGTC ATGAAGCGAA AGACCGTTCT 6420

TCTTAGCCAA TTCTTCGTGG ATAAGGATTT TCTTGGAATC CCCTTTTTGA AGGTGTCGCC 6480

CTTCTTTTAG ATTGAAAGCC GAACTGGTAA AGGTTACATC CTTGGATGAA TCCTCAAGAG 6540

CCGTTAAGCT AACCAAGTTA TTGTCTGCAG CTGATAAATC ATCACGCTCC ACGCTCTGCT 6600

CGCCAGTCAC TGCTTCCTTG TCTTTTAGTT TTGCGACCGT CTCAAGTTCA GGAGAGACAT 6660

TTTCCAGCCC CTTAATCTTG CTTACAGATG CTAGGTCTGA CAACTTGAAT GTCTGACCAT 6720

TCTCTATCTT CTTAATAGAA AAAGATGTAT TGAGTGATTT ATAAAGATTG CTTTCTAC"G 6780

TTTTGTTGGA CTTCATCAGA GTCAAACAGG CTGAAATTCC GGCCAATAAG ACCAATAAAA 6840

TCAGAAATAA AATAAAACTT CTCAGTCGCT TTCTGCTGAC ATAAGCCCAA GATCTTTGGA 6900

TTGGATTCAT TTGTCACCTC CATATTTGTA AGACTATTAT AAAACCCAAA TATGAAATAT 6960

TTATGAAATA CGAAAAAAAA ATATCGAGTA GGGGATAATC TCTAGCCCCT CTCACACCAC 7020

CATACGTGCC GTTCGGCATA CGGCGGTTCA ACTAACTTTT AACGCATGTC GTTCAAGGTA 7080

ATAATCCAAA CACGAAACCA GTCCACGTTT TTCAAGGACT GGTTTTGATA TAGCACGTTT 7140

AAGTACCGAC TTCTGAGCTA CTATAGTAGA TTGAAACTAG AATAGTACAC CTCTACTTCT 7200

TCTGAATGGA AATGTCTAGG CTCGACAAGG TCTTGCTAAA GCCATCTACT TCTGCACCAC 3720

CACCAAGGTT AGAGGCTTGA GCCGCCTTAC TAGCCTGTTT GGCAACACCT GAAGTCACAT 3780

TGGCAAGGAC AGTGTTTCCA ATTGCACGGG CAGTGTAATT AGCTAGGAAG TAAGCAGAAA 3840

CTAGAGCAGG GATAGCAATC AAGATAGATT CGGTGATGAA TTGACCCAAG ATACTTGCCT 3900

GCTTGAGGCC GATAGAGAGG AGAATTCCCA CTTCCTTGCG ACGGGCGTTG ATCCAAAGGC 3960

TGAGCAAGAG GGCAAGGAGG AGAACTGAGA AGCTCAAGCT ACCCCAGAAG AGGAGGTTGG 4020

CCATCTTGTA CATACCAGAG ATAGATTGCT CAAGAGCTGG GTAGTTAGAG GAGCTCTTGA 4080

CGAGTGTGTA GCTCTTCCAG TTGATACCAC TGATGCCATT CAACTCTTTC ATAACATCAT 4140

CCAAGTTCTT GTCTGCTGTT ACAAAGAAGG TTGCGTCCCC ATAAATGGCT GTGTCTTCTG 4200

TGTATCCATA AAGTTTTGCA GCAGTGTGAA TGTCTGTAAT AGCTGTGTTT TCGTAAAGTT 4260

CTTGTGAGTA GGTTACTGCT GACTTATTAT GACCATCAAA GAGTCCCTTG ATTGTCACTT 4320

CAACTGTTTC CTTGGCTCCT TTTTCATTAT CTGCATCGTA GATATTAGAG TCCAGTTTAA 4380

CCTTGTCCCC TACTTTCCAG CCGTGTTTGG CTGCCAAGTC CTTGTGCAAG AGGATTTTAT 4440

CCTTGTCGTC GTTGGTTAAG TGCTCTCCTT CGACTAGTTT ATAAGAACCA GAGACAAACT 4500

TGTCTTCTTT AGAGGAGTCA TTGACACCTG TAATCATCAA GCTACTTCCA AAACGCTTGG 4560

CACGATCAGC AGTGAGATTC TTCTTGGTTT CTGGCGTTTC AATCAGGTCA TATCCAGTCA 4620

AATCTCCGAT AGCGTTGATA CGTTTGACAT AAGACTCAAT GGCCTTGTTT TCGGTGATTT 4680

TTTTGATGTC TTCACCCTTG ATATTCCCAG CACCACGAGG CGTTCCTTGG TTGACGCGAC 4740

GATTGATTTG CATGGAGAAG CTATTGGTGA TATTTTTAAA GGTCTCCTGA GAAGCCTTGG 4800

CAGTAGCTCC CTTGATTGAC AAGCCGACCA AACTCAAGCT CGCCATGAGG AGAATAATCA 4860

GGAAGATGAC AATCGATTTG AAAAACTTCC TTGTAACATA GGCAAATGCG TTGTGTAACA 4920

TAGATTCCCT TTCTAGATTT TGTTTTAATC ATTCTATTAA AATAAGCTCA AATTATTTAC 4980

TAGTATTGCG CGTTTCAGTC AGTTTCTTAT CCTTTAATTC AAGTGTAATA TCTGACGCTT 5040

GTGCCACTTC TTTACTGTGA GTTACGACAA TCACACATTT ACCTGTTTTC TGGGCAAGTG 5100

ATTTGAGTAG TTCGACAATA TCTCCAGCAG TTTTAGGATC CAGATTTCCT GTTGGCTCAT 5160

CAGCTAGAAT AACTGGAGCT TCTGAGACCA AACTGCGAGC AATGGCAACA CGTTGCTGTT 5220

GACCACCTGA TAACTGGAGA ACATTCCGCT TGATCTGGCT TTCATCCAAA CCAAGCTCAA 5280

GAAGTGTATT CTTGCTTGCC TTTTTGTTGA CCAATCGGAT ATTTTCCAGC GGAGAAAGAT 5340

AATCTATCAA GTTATAATTT TGAAAGACCA GGGAAATATG GTGCATGCGA TGGTAAGAAT 5400

AGCCCTTCTT ACGAATATCC TCTCCTTGAA AAAGGATAGA ACCTTCAACA GGACTATCTA 5460

ATGCAAGTCC GCAATAACAG TCAAGAGATG CTGGTAGAGG CTATTGATTT GTTCCTTGAG 1980

ATTACCTATC TCATCCTTAG AATCCACGCG CAATCGCACT TGGGAATCCA GGTCCATCAT 2040

CCGACGGGTC ACCCGCTTGA TTTCCAAAAT CGGTGCAACA ATAGTCCGAG CGTAGATGTA 2100

GGCCACCAAA AGGGAAATCA GAAAGGAGGC CAGCAAGGTA TAGGGAAGAA ACTGGAGACT 2160

GATTTGCTCC GCTTCCTTTT GTAAATCCAT GGAAGCTAGA AACTGGAGAA TCATAGTACC 2220

ACCGTCTTGC GTTTTCACCT CGCGCTCCTC AATAAAGAGA GAGGTTGTCT GGCGGTCTGT 2280

GTCCAGAGGA AGACTGTCCT TGACTTCTAA CTTGTCCTCG GTCATCTCAC CTTTGACGGT 2340

CCCCTTGATA TCACTAGTCT GGGAATACAA GTCTAACACT TGCTCGATAC TCTGCCTATC 2400

TTTCCCTTCT AGGGACTGGG CAATGGCTGT TGCCTTTTGA CCAATGGTTT CCTGACGATG 2460

ACTCAGATAA GTCGAAGGAA AAAGAAAATA AATAGCTAAA TGAAGGCAGA TAACCAGAAC 2520

ACTAAATATC GAGAAGGTAT AGATAAATAT CTTTGCAAAT AAACCTGTTC GTTTCATTTT 2580

CGCTCCAATT TATAACCAAC ATTGCGCACA GTGAGGATAC AATCCAAGTC TAGCTTTTTC 2640

CGCAATTCCT TGATATAAAC ATCAATAACA CGGTCAAAGG GAACCTCATC TGTCGCTTTC 2700

CAGACGGCAT CGATAATCTG AGATCGAGTC AAGGCCCGGC CTTCATTTTT CACTAGATAG 2760

TCCAGAATTT CCAACTCTTT GGCATTGATA GGCACTTCTT GACCTGCGAG GCTTGCACTG 2820

TAGCTTTCAA AGTCCACCTT GGTATCCTTG TAAGAAAAGA TTCGTCCTGT ATCGTAGTAG 2880

CGCTTGAAAA TCGCGTCCAC CCTCACTTTT AAAAGGGAGA GGGAGAAAGG TTTTTCCAGA 2940

TAGCCATCTG CCAAAGAGGC AAAGGCACTC ATCTTGTATT CCTCATCTTG AAAAGCTGTC 3000

AACATCAAGA CAGGAACCTG ACTGGTTTTA CGAATCTCAG CTAGGACTTC TAAGCCGTTG 3060

AGCTTGGGCA TCTGGATATC CAGTAAAACC AGGGCCACCT CATAGCTAGA AAATTGCTCC 3120

AGAGCTTCCT GACCGTCCGC TGCCTCAATA GTTTCATAGC CACAATCCGT CAAATAATCA 3180

CTGACCCCCT CACGGATCAT CTCTTCATCT TCTACAATTA AAATTTTCAT ACTTTAACTG 3240

CTCTCTATTT TTTATTTTTC TTAGAATAAA TACCTACCCT ATTTTCTATT ATAGTCTCTT 3300

GCTGGCCTTT TGTCTGCAAG CAACTGACCA CTAGATAAAA CGTTGTGAAA TTCCTTTCTC 3360

ATAAATTCCA TAACTTTAGT ATATTATATT TAAGCACTAA AGTACAAAGA AAGCAACTGA 3420

AAGCAATGAT TTTCACCACT GCTTTCGGAT TTATTTTGAA TTGTTAAATA GCCATTCCTA 3480

TCCACTATTC TTGAATAGAA ACACAAGATG CAATCTTTAT TCTAGACTCA TTTTTTCAAA 3540

TTTATTCACC ATCCAGCAAG AGCTCTTTTG GTTGTTTTCT AAGGAGATTG CTTGAAGCAA 3600

GCGCCATAAC GAGAACCACT AGAACCAAGG CAAGGACAAA AATGATGATA AAGTCTGATG 3660

940

AGCAGCATTG TGCTAGTCkA GATTCAGTTT ACTATACTAA AACGAGTAGC TTGAAATCAA 180

AAAACCCACC CTCACAGGCA GGTTTTATCT GTATTATTCA GCTAGATTAT GCTTTACCTT 240

CTGAACCGAA TACGTCGATA CGTTCTTCAA CCGATGCTTG GATAGCTTTT ACACCGTCAG 300

CCAAGAATTT ACGTGGGTCG AAGAGTTTTT TCTTGTCGTA TTCTGCTTCG TTTGCTTCGT 360

AGTCACGAGC AAATTTACGA GTTGCGTTAG CGAATGCGAT TTGGCATTCT GTGTTAACGT 420

TAACTTTGGC AACACCAAGT TTGATAGCTG CTTGGATTTG CTCATCAGGA ATACCTGATC 480

CACCGTGCAA TACGATTGGG AATCCTGGAA GAGCTTCTGT CAATTTTTGC AAGTGGTCAA 540

GGTCAAGACC TTCCCAGTTT ACTGGGTAAG GACCGTGGAT GTTACCGATA CCAGCTGCCA 600

AGAAGTCGAT ACCAGTTTCA ACCATTGCTT TAGCGTCTTC GATTGGAGCC AATTCACCTT 660

TACCGATGAT TCCATCTTCT TCACCACCGA TAGTACCAAC TTCAGCTTCT ACTGAGATAC 720

CTTTAGCGTG TGCTTTTTCA ACAACTTCTT TAGCCAATTT AAGGTTTTCT TCAACTGGAA 780

GGTGTGAACC GTCAAACATG ATTGAAGTAT AACCAACTTC GATACACTCA AGTGCATCTT 840

CGTAGTGACC GTGGTCAAGG TGGATAGCTA CTGGTACAGT GATACCCATT GATTCAACAA 900

GGTTAGCGAT CAAGTTGCGA GCAACTTTGT AACCACCCAT GTATTTAGCA GCACCCATTG 960

AAGTTTGGAT CAAAACTGGA GCTTTTTTAG CTTCTGCTGC GCGCAAGATA GCTTGAGTCC 1020

ACTCAAGGTT GTTTGTGTTA AATCCACCAA CTGCATAACC GTTGTCACGG GCTGCTTGGA 1080

CAAATTTTTC TGCTGAAACG ATTGCCATTT TATCAGGCCT CCTGTATATT TTTATGGGTC 1140

ATCCCATTTA CATTGTTCAT TTTATCACTT TTTGCCAAAA AAATCTAGTT TTTCCCGCAG 1200

TTTCGATTGA TTTTCTTCTA ACTCCATCTA TGTAAACCCT TTCTCTCCCT AGTCTTGGAC 1260

GACTTTTGGA AAATCTATAA AGAAGGTTAA ACTATTCTCC TCCATCTCGA AACGATAAGC 1320

TAATTTTTCA TGTTCTAATA GACTCTTAAC CACAAAGAGC CCCATACCAG ACCCCTTGAC 1380

CTTGCGACTG GCATTGTCAG AAAAAGACTG GGCTAGTTTT TCTTGTTCCT CTGAGCTACA 1440

GCTATTTTCG ATAAAAAGTT CTCCTTCTCT TTCTCCAATT CGAACTAAGC CACCTGGAAC 1500

AGAGTGCTTA ATGGCATTGC TGATGAGATT AGAAAGAATC AACTTCATAA CTGATGGGTT 1560

TAGATAAGCC TGCTGATGGG TCAAACTATT GTCTATCTGG AGCTCTCTTT CCTTGGCTAG 1620

CAAGGCATAA TCTTTGACCA GATTTTGCGT CATCTGGAGG AGGTCAATTG TTTCCCTATC 1680

ATCTCGCAAT TCCTGCACAG AAGAGAGGGA AAGTATCTGC AGAACATGGT GATTGAGTTC 1740

ATCCACAATC CCCAAGGCAA CTCCCAGATA CTGGTCTCTA TCCTTATAAC GACCGATATT 1800

CTCTCTCATA TTTTCGATTA GGATTTTCAA ACTAGCCAGC GGTGTTTTCA ATTCATGAGA 1860

AGCTCCTCGT AGGAATTCGA CCTTCATCTT CTCCAGCTGG AGAATGGCTT CATTCTTTTC 1920

GATGAGCAAG AAGTATAGTG TCCAAGCAGA ACCCCAAGTG ATTGTAGCAA GTGGTGCCCA 14220

ACCAACGTCG GTAATACTCA ATTGGATACC AGTGTTTTCA ACGAATTTTG CTAGTGATGC 14280

TGAGAAAGCA GTGTTTAGCA TACCGATGAT AGCACCGATA CCTGTAAGAG CGATGGCAAG 14340

TTTGATACCA CCTTCAAGCG CTTTGGAGAA TTTCACTCCA AAAAGTAAAG CCAATACTGT 14400

CAAAATGATT AACATGATGA CAGGTCCACC CATTTCTAAG ATGGGATTGA AAACCTTTCC 14460

GATTAGGTCA AAGATTGCAT CCATAACAGT TCCTCCCTTT TTGATGTTAT ATGAATGTTA 14520

ACAAATTAGA ATTAGCTTAA TCCGTGTTCT TTAATAGCTG CTTCAATATT GTCAAATACT 14580

GGAGCGCTCA TTGCTGGGAT ACGGAATAAG ATTGGCCCAG CTTCGATAAC TGGGATACCT 14640

GGTTCAAAAC CAAGGTCTGT TGCAGCGATT GGTGTAAAGA TATCGTAACC TTTCATAAGG 14700

TCTTCGTTTA CATCTTTCAC CATGACTGCA TCACAGTGAA CATCATAACC ACGGTTTGAA 14760

AGTTCTTCTT CTAGAGCACT TTTAATTTGG TGACTTGAGT TAACACCTGC ACCGCAGGCA 14820

GCAAGAATTT TAATCATTTA GATTTCCTCC GATTTTATTT TTTAATAGAC AAGATTAAGC 14880

GGTTGCTTCA GCAATGTAAG TATAAAGGGC TTCTGGTTCA GAAATTTTTG ATAGGTCTTC 14940

AAGATGACCA TTTCCTGTGA AGAAGTCCAT TAACTGAGCA AGAATGTTCG TTTGACTTGA 15000

ACTTGAATTA TTAATGATAA AGAAGAGTAG GGATACTTCT ACTTCCTTAT CAGGAGCTAT 15060

CATATTGTGA AAAGTTATTG GTTTTTCTAA TCGAACAACC ACCACTTTCT CAGCTAGATT 15120

ATGAACAATA TCTGTGTGAG GAATCGCTAC ATTTGGCAAG TCCTTTCCTA GAAATTCCAT 15180

ATCTAAACCA GTTGGAAATG ACTTTTCACG CGTGATCAAG GCTTCACGAT AAGTTGGAGT 15240

GACAATTTCT CGTTCTTCCA ATAAAGTTGC AACCTGATCA AAGAGTTGTT CTTGACTATC 15300

CGCTTCTAAG CAAAACACAA GGTTTTTGTC AAAGAAATAA TCTAATACCA TAAGTTTTTC 15360

CGG 15363

(2) INFORMATION FOR SEQ ID NO: 140:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28882 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

TAAGACTATT TAATAGTGGA GTGAAATAGG ATACGAACAA ATTGATTAGG AAAATCAAAT 60

GAATTTATAG AAATCTTTTA GCAGTTATGT TATCCTATTC TAGTTTCAAA ACGCTATAGA 120

TCACTTTGTT CTGTCCTTTC AACTATATTT TTAGTTTCCA AAGCTTTATC AGCCTTTTCT 12420

TCTACTATCA TTTTTTCCTC TTTAGGTTTC TCAGCAGTAT GAGTAATAAG TGTTTCATCC 12480

GCATAAACTA CAGATTCTCC AGCTATATTT CCTCCTAATA AAACTGCACA AGTCCCAATC 12540

ATTACTGAGC AAGCTCCCAC AGCAAACTTA CGAATGCTAT AAACTCTTTT CCGATTCCAA 12600

TGGCCTTTCC CCATAAAACC CTCCTTATAT TATATTTAGT GCAGTTAGCT ACTACCAAAG 12660

CCCAAGTGGT ATACATGGTA TGACAACCTA GTTTCAACAA TTTACACTCT GCGAAAATCC 12720

AATTCAAACT TCGTCAGTGT CGCCTTGCCG TAGATATGAT TACTGACTTC GTCAGTTTCA 12780

TCTACAACCT CAAAACCATG TTTTGAGCTG ACTTCGTCAG TTTCATCTAC AACCTCAAAA 12840

CCATGTTTTG AGCTGACTTC GTCAGTTTCA TCTACAACCT CAAAACCATG TTTTGAGCTG 12900

ACTTCGTCAG TCTTATCTAC AACCTCAAAA CTGTGTTTTG AGCAACCTGC GGCTAGCTTC 12960

CTAGTTTGCT CTTTGATTTT CATTGAGTTT ATATTTTATA GGAGCGCATT ATTTTGCTTT 13020

TGCTGCGTAC TCTTCGTTAC GTTTGATCAT TTGTTTTCTG TACCAAGCAA AGATACCGAT 13080

ATAGAATACA AGGAAGACTA CTGCACCAAG GATTGCTTTG ATATCACCAG TTGTAGTGTT 13140

ACCAATTGTC CAACCAAGAA GTTTTTCGAT TGGTCCTTCA AGAGTAGAGT GAGTAATCAA 13200

TTGAGTTTGG CTCACACCTT CTGGGAAGGC ACCTACACCT TTAGCAAGTT CTGTTGCAAA 13260

TGGTGCAATA AGTGTACCTG AAAGAAGGAA GAGTGGCAAC AAGAGTGTTC CGAAGATAAT 13320

CATACGGAGC AATTTACCAC GAGTTACAAC CAAGAGAGCT GGAGTAACAC CCATAGCGAT 13380

GATACCTGCA AGTGGCAAGA TACCATTTCC AACTTTTGAA AGAAGCACTG CTTCAATCAA 13440

CATGATTGGT GCAAGTACGT TGGCACAAGC CCAGATTTCA GCACGACCAG CGATGAATGG 13500

CCAGTCAAGA CCGATATTGA ATTTACGTCC TTGAAGACGT TTAGTAGCAA CGTTTGTAAT 13560

ACCTTGTGAT AGTGGTTCTA CGGCTGCGAT GAACCATGAA CCGATAAGTG AGAAGAGTTC 13620

CAAAGATACA CCGGCAGTCA AACCAAGAGA CAACCATCCT TTGATAACAA GACGCCATTT 13680

ATCTGCATCT GCAACACCTG CAATTGGATG TGGAGTTCCC ATAATACCGA TAACGATACC 13740

AAGGATGAAA CCGATGAAGA ATTTAGATCC CCAGAAACCG ATTTTCTTGT TCAATTTAGC 13800

AGCATCAAAG TCATATTTAT CAAGGCCTGG GAAGAATTTT TCAAAAATCT TATCCAAAAC 13860

CATGATAACT GGGTTCATCA TGTAGTTCAT GTGAGTTGAT GTCATTGGTG ATGAACTTGG 13920

GGCGTTAAGA AGGTCATCAA ATGTAGGTTT CATCAAGTCA GAGTTGATAA TTTTCAACAC 13980

ACCGACAAGG ACGATAGCTG CTGTAGCAAT AAAGAGTGAA ACCCCTTGAC TCACACCATT 14040

GTTATCAGCA TACCATTTAA TCAAGAGACC TGTGATAGAC AAGTGCCAGA TATCAAAGAT 14100

ATCGACATCA AGTGTATCTG TTTTCTTCAT AGCTAGCATC ACTATGTTGA CAATCAACAT 14160

TGTTTCAATT CACGTTCAGG GCGATAGTAA CTTCCACGTG TACGGGTAGC TGAAGATGTT 10680

TCTGATCCAT AAATCAACCA TTTTGGATGC TTAGCTCTAA GGGCTTTGTA ATTATCTTCA 10740

GAATAGTTAA ATCCAACAGC ATCGAGTTCA TCAGCAATTT TCTCATGCCC TCCGCTACCA 10800

TTACCGAAAC GGAATTTATC TGCTCCCATG GTAACATAGC GAGTCTTATC AACATCCTTG 10860

ATAACCTTAA CCAAACGTTT AACAGTTGCT AAAGAGTGGG CATCACCATT AGCTTCACCT 10920

ATTTCATTAC CAATTGACCA CATGAAGATA GCAGGGTTGT TTTTGCCTCT TTCGACCATG 10980

GTACGTAGGT CAAAATCAGA CCATTTTTCA CCTTTTCGAG CTTCTGGGTG AGTGGCATCT 11040

TTTTCAAAGA AACGTCCATA GTCATAAGGT TTCTTGCCAC CATACCACGT ATCAAAGGCC 11100

TCTTCCTGAA CGAGTAAACC TAGTTCTGCT GCGATTTGCA AGGTTTGCTC ACTAGCAGGG 11160

TTGTGGGTTG TACGGATGGA GTTAACTCCC ATCTCCTTCA TTTGTTTGAG ACGGCGATAT 11220

TCTGCTTTAT AGTTTTCTTC TGCTCCAAGC GCCCCATGGT CGTGGTGCAA GGATACTCCA 11280

TGGAATTTAA TACGTTCACC ATTCAAAGAG AAACCTTCAT TTGGAGTCCA GTGATAGTAA 11340

CGGTAACCAA ACAAATCCTT CTTAGCATCA ACCAATTGAC CGTCACGGTA AACACGCGTA 11400

ATCAATTCGT ACAAGGCAGG TTTGTCATTT AAAACAGTCC AGAGTTTTGG TCTTTCAACT 11460

TCTAAAATCG CATCTAGGCT TGTTGATTCA TGTGCTTTTA AGGTACGACT CGCTGTACGA 11520

ACTAAGCCTG TTACAGCATG ACCACCTCGT TCAACGATTT GATATTCGGC TACAAGTTCA 11580

TGGTCTTTGT CGTCCGTATT GACGATTTTG CTGGTCACAT GAGTTTCAAC CTTGCCATGT 11640

TGTTGTTCTT CAAGTTTTGG TGTTAAAATA GTTGTCCCAT TTTTCTCAAC ATGCACCTTA 11700

TCTGTCACTT GTAAAGTCAC ATCACGATAG ATACCACTTC CTGAATACCA ACGGCTACTT 11760

GGCTGTTTGT TGACTGCATG GACAGCAATC ACATTCTCAC GACCATCTTT TTGAAGGTAT 11820

TTGGTGATAT CATATGAGAA CTGGTTATAA CCATTTGGAT AATGCCCCAC TAACTGACCA 11880

TTGACATAAA CTTGAGAATC CATGTAGACG CCATCAAAAG TAAGGCGAAC ATTTTTCTTG 11940

AGGTCTTTTT CATCTAGTTT GAAAGTCTTG CGATACCAAG CTTCCCCACC GTTGAGCTGT 12000

CCACCTTCAT TTTGTGCAGG AGATTCATGA TCGAAATCGT TAAAGATACT CCAGTCATAC 12060

GGTAAATCTA ATTTTTTCCA CGTAGATACG TCTGCATCAG GTTTAATGGC TTCCTTAGAA 12120

TTTGCATTGA GTTTAAAGTA CCAATTTTGA TTAAAATCCA CTTTCCTGTC TTCAATCATT 12180

TGATTCACTT CTTCATTTGT TACAGCTTTA GCATCTTCCT TGAGCGGTTT TTCTTGATTT 12240

GAAGCTTGTG ATTCTATCCT TGGAGCTTTT TCTTCCGGTT TAGCAGACAC TTTTTCCTCT 12300

TTTGGAGTTA CGGCTTCATC TTCTTTCTTC TCAGATGCAA TAGCCTCAGT TGAACTAGGT 12360

GTTACTTTAT CAGCTGGTAA TACAGCTTGC GTTCCATCTT GATAGTGAGC TCGAACCGAC 8880

AATTTGACAG TTTGGTCTTC TTTGAGACTG TCAGCTTTTT CCACTTGCAA GCTCAAGTGA 8940

GCAATTTTTG GCGCTTCTTC AAGGAATTGA ATTGCATAGG TTTGAAGAGG GCCACCATCT 9000

TTAGGCTGAA TAAAGATGCT CGCACGCATG CCGTTTGCTG CGCTTGCTTG AAGAACTGTA 9060

ACAGCTGCAT TTTTAGCACT TGCTGTGACT TCTGGCAACT TAGCTCCATA AGCAAGAGTG 9120

CGGTATTGCA TTGGTTTTTG ACTAGTAAGA CCTGTTACTG CCTCACCACC AACCGTTACA 9180

GTTGGTACTG CAGGTGCCGC AGGATTGCCT TCTTCTACCA CAAGGGTTGC ATGAATTGGT 9240

TGACCTTCTA AATAACCGGT CGCTTGAATA CGAGAACCTG GAATTGCTAA CTTAGCTTTA 9300

TCTTCTTCGG CAATCTCCCA CTTGTCCACT TCATACTCTT CAACACTTCC ATCAATCAAA 9360

ACATAGGAAA CAGATTTGTC TACAGAATTC AAGTCAGTAT TTGGAGCAAT ACGTTTCACA 9420

ACTGGTAGCT CTGATTTAAG AGCAATCACT TCTACACGAG CTTCTACTTC TCGTCCGTCA 9480

GCCATACCTT TCACCGTTAC AATACCAGGC TTGCTCACAT CTACTGAAGA CCAGGTTACA 9540

GGACGTTCTG CACGGCTACC ATCACTGTAT ACAAACGGAA CAGTGGTAGG CATTTCAGGT 9600

GCCTCTCCAA TAATGGTCTG TACTTTTGGC ACTTCTGTCC CCAAAACAGT CTTCTCTTGT 9660

CCTTCTTTCT TACCAGTAAA GACAGTGACT TGGTTCGATT TCAAGAGATC AGAGTGGGCA 9720

GTCAGGGTGA ATTTCCCTGC TTGTTCAGTT GATTTGACAA TGGCAACACC TTTACCATTA 9780

AATGCTTTAC GAATCCAAGA ACCATCTGCT TGCGCCTTAT AGCGTTCACG GCTGGCTTGT 9840

TCTCCGTTAT CTACACCGAC CAGTTGACCT TGGCCATGCA ATTGGAAGCG AACCAGATTA 9900

TTAGCAGTTG GAACCACATT CCCCTGGCTG TCAACAATTT CATAGTAGAT GTAAGTCAAG 9960

TCTTTTCCAT CTGCTGCAAT CGCATGGTCT TCCTTAATAA GACGAACTGC CGCTGGCTTA 10020

CCAGCAGTCG TAATCTTATC TCGAGCAATT TCCTTGCCAG ATTCATCACG AGCAATTGCT 10080

TCCAAGGTAC CTGGTTGATA GGCAACTTTC CATTCAAGAT AAAGTTCATT AGCATTTGCA 10140

CCTTCTTGGT AAGTCCGCCC ATCGCTGGTT TGTTTTTTAT TGAAAGTCTT AAGACCAAGA 10200

GATTTTCCAT TCAAGAACAA TTCTACACTA GAAGCATTCG AATAAGCACG AACTGGAATC 10260

TTACCTTCTG AGTCAGCTAC TTTGGATGCT AATTCTTTGT TTTCCCAGTT CCAGTGAGGA 10320

AGAAGGTGTA CCATCGGTTT CTTCTTAACA GAAACCCATT GGCTTTGGTA GAGATAGAAG 10380

TCATGTTTTG GAATGCCGGC TGTATCTACG ATACCAAAGT AAGAGCTCTT AACAGGAGTT 10440

TGATTTTGGT TGTGCCATGG TGTAGGTTCA CCAATATAGT CCGTACCTGT CCAGATAAAC 10500

TGTCCAGCAT AGCCAGCGTT GTCACGGTCA AAAGTCCATG AAGCGGTTGC TGTTTTCCCC 10560

CAACCCACAC GATCATTTCC ATAATCTGAC TGTTCATAAT TACGCTCAGG TCCATTGCTA 10620

TCATGAGAGT TTTTGTCAAT ATCATTGGTT GCTGAAGCAA AGGCCTGGTT ACTGTTTTCA 7140

TCATAGTTAG GGTTATCTGA AAGAGTCTCA CCAAGTTTGT CTGTCACTCG TACAGTGATC 7200

TCAGCAACAA GGTTACTACC AAGGACACGG CCTCGAACAG TAAATTGACC TGCTTTTGTC 7260

AGATTTTCCG CTGGAACTTC TTCCCATTCA ACTGTCAGGT CTTTTGTTTC GTAGCCGTCT 7320

TTACCTGTGA AGTAAACTGG AACCTTAGTC GGCAATTCAA GTGCTTGACC TACTTGTAGC 7380

AAGCGAGCTT GTTTAACCGC AGCAACTGGT TTATGAGAAA GTAAGCTCTT ATCCTTAGTG 7440

AAGTGCAGAC GGTATTCTCC TAAGATGTCG CCATTTTCAG CTTTCGCGAT GACACGAACT 7500

GGCTCACCTT CACGAACGCT TGGAACGACG GTAGCGAGAC CATTGTTGCT AACACTTGCT 7560

GTGACTGCCG GAACTTTTCC ATCTACAGAC TCAAGGTAGT AGTCTGTCAA ATCAGGGTTG 7620

AAGTTTGCTA AGTCTTTGCC GTCAACTTGG ATTCTTGTTT GTCCTTGCTT GGCTGCCGCA 7680

ACTTGTTTCG CAAAGATTTG TACCTCTGTG ATAGACGTTC CACGCTTGTT ATCTGCTTTA 7740

ACCATGCGAA TACGAACAGC ATAGGTTTCA ACTTTATCAA AGCTAAAGTG GTTCATTTCT 7800

CCAGCCTTGA GTTGAGCAGG GGCTTTTAGA TTAGTAACTG GTTTCCAGTT GGCAGAATCA 7860

TTAAAGACAT GGTCCTCATT ACCAACAAAA CTAGGGTTTT TAGGAGCTGT TGGGACAGTC 7920

TTACCAACAT AATACTCAAT CACATAAGAC TTCGGTACAC CAACTCCATG GTCTTCATGG 7980

AATCCGACAC TTAGATTATC AACGGAGCGT TTGCTCAAGA TACCTGAATC TCCAAACAGA 8040

ACACCGACTG AAGCTTCTGG ATTAGTACGA TTCCAGTTTG TCCAACGATT GGCTGGTTGG 8100

TTATTGTAGG AAATGAGCTT GTCATTAACA TTTGAAACTG GGTCGCTTGG ATTTGAGTCT 8160

GAAGCAAAGG CAAGTGGCAA TTCTGAACCG GTCCATTGGT CAGAAATGTT TGCACCTTGC 8220

TCAGTTTGAG CAGATACGCG AACATGAAGT TTAGTTGTTA ATTGCGTACC TTCTAAGCGA 8280

CCATTAACTG TAAAGACACC TTCCTTAGCG TATTGCTCTG GACGAATCGC ATCCCATGCA 8340

ACCTTAGCTG ATGAAACGTG ACCATTTGAA TCATATGTCC GAACACTTTC TGGTAATTGT 8400

GGTGCTTCTG CGATTGGAGT TGTCACACTG ACTTCTTCAA CTGAAACGAT ACCTTCTACA 8460

GAGACTTTTG CACGCGCTTC AAGGTCAATT CCTTCAACTT TACCTAGTAC TTCAAATGTT 8520

TGATAGGAGT CTAGTTTTTC TTTCGGAATA GCTTGCCAAG TGACTTTATG AGTTTTAGGG 8580

AAACCTTTGT CATACTCAAC TGTTACTGTT GCTGGAAGAC TTGGTTCCTG ATGCAAATCT 8640

GTCACTACAT TTACAGGACG GATGGATTGC GCAATCTTCT TCTCAGTATT GGCTTGGATA 8700

GTGAGTTCAA CTTGGTCTTT AGCTCCCTCA TATTCAGCGT TCAGAGTGAC TGCTCCTGGC 8760

TTATGCAACT CAAGCATTCC TTTACGAATT GCGACTTCCC CTTCACCACT TGTAGAGAAG 8820

CTATAATGAC AAGGAGGATA GCTACACACA TCCTGATGGC TGGTATTATC GTTTTCACCA 5340

TACCAAATAT CAGAAAACAC AGACAGACTT TCAACAAGAA ATCAAGGTTT ACTACGCCGA 5400

CGAACCTGAA TCAGCCCCTC AAAAGGGACT GTATATGAAC GAACGCTATC AAAACTTGAA 5460

AGCTAAAGAA TGTCAGGCGC TTTTATCTCC CCAAGGTAGA CAGATTTTCG CTCAACGCAA 5520

GATTGATGTG GAACCTGTCT TTGGGCAGAT AAAGGCTTCT TTGGGTTACA AGAGATGTAA 5580

TCTGAGAGGG AAGCGTCAAG TGAGAATTGA CATGGGATTG GTACTTATGG CCAATAACCT 5640

CCTAAAATAT AGTAAAATGA AATAAGAACA GGACAAATCG ATAAGGACAA TCAAATCGAT 5700

TTCTAACAAT GTTTTAGAAG TAAAAGTGTA CTATTCTAGT TTCAATCTAC TATACAATAA 5760

GAGAATGACT CAAAATTAAA AAGCTAGAGT TCCACAATTG GAAATATCTA GCTTTTTTGT 5820

GGTTGAGAAC TATTTTGTCT CAGGCTCTTT ATCTTCTATT TAGGACAAGA GTTTTTCTTT 5880

GGTCTTTAAT GATAAAGAAG GTATCAAAAT TTCTAGTCTT CTTTTTTACC TTTAGTAACT 5940

ACTAATCCTG CACTCAAACC TAGAAGAGTT AAACCTGCTG CTACTGCTGC TTGGCTTGCC 6000

GCACTACCTG TACTTGGTAA CTGGGCTTTA TTAGTTTGAC TAGCTTCACT TGAATCAATT 6060

GGTTTTGTAT CTGCTTTTTC TGACACTTGT GGTTTTTTAG CTTCTTGAGC TACTGGTTTG 6120

GTTCCAACCA AGACGATGCG GTCTGTCGGA ACTTCTACCA CTTCACGGAG TTTTTCTTCC 6180

TTACTTCCAT CAGGATTAAT CGCTGTAAAG ATACGTTCTT TTCCAACTTT TCCTTCTTGT 6240

TCTACACGAG TTTCACCTAG ATACAGTGTT GAATCTTTTT TCTCAACTGT CTTGTATGCC 6300

AAATCTTTTT CAACAAATTC GATTTTTGGA AGATCTTCTT GTACAGCAGC AACTGTCTTC 6360

TCAGAAACTG GTTTTTCCTT AGTCAAGTGG ATACGGTATT CCTTGACTTG TTTTCCACTT 6420

TCTGAAACGA GGCGAACAAG TACTGGAAAG CTATCTTCTC CACTATCTAC CACAGTTGAA 6480

GCTACTTGAT TGTTTTCTTC AACTGAGACT TTTGGCCGTT GACCTTTATA GGTAATTTGA 6540

TAGTCTTGAC GATTTTCAGC GAAATCAGCA AGTTCTTTTC CATCTACAAG AATCTTTGAT 6600

TGAGTGCTTT CTTGAGGCAA TTCACTTGGT GCAAGGAAGG TCATCTCAAT CATCGCAACA 6660

CCGCTCTTAT CTGCTTTACG CTCCATACGC CATCTCATAG CTTTGGCTTT GATAGCTTTA 6720

AATGTTACGT TGATTTCATC ACCAGCTGCA ATGTCTTTAT CCGCACGATA AGGAACAGCT 6780

TCCCAATTTT CTGGATTGTT GAATGGATGG TCTGCGTCGT AGGCTTGGTA GTTTGAATAG 6840

TAGGTTGGCA CTTCAAACTC TGGACCGACA TAGCGTTCTA AAACGAGTTT AGATGGTGCA 6900

TCCGTACCAC TATCTGCAAA GAACTGAACT TTTCCTTGTG TAACAGTCCG TTCTACAATC 6960

TTACCATTTT CACGGAAAAT CACACCCGCT GATACTTCTG GATTAGAAGA TGGTGTTGGT 7020

GACCAGTTTG TCCAACGACG ATTTTCTGAA TGATCTCCGT CATTGAGATA GTCAACGCGG 7080

TCAACCATCA TCTCATGGGT TTCACCAGGC AAACCATTGA TCAAATGGGA AACAATCTCA 3600

ATTTTTGGAT ACTTTCTCAA ACGCTTGACC GTTTCCACCT ACAATTCATA AGAATGCGCA 3660

CGGTTAATCA GGTCAGAGGT TGCTTCATAA GTAGTTTGCA AGCCCAATTC AACCGTCACA 3720

TGCATGCACT CCGATAACTC AGCCAAATAT TCGATGGTTT CGTCTGGTAA ACAGTCTGGG 3780

CGCGTTCCAA TATTGATTCC TACCACACCT GGCTCATTGA TAGCCTGTTC ATAACGCTCT 3840

CGAATAACTT CCACCTTTTC ATGGGTGTTG GTAAAATTTT GAAAATAAAC CAGATACTTC 3900

CGAACATCCG GCCACTTGCG GTGCATAAAG TCAATTTCCT TATAAAATTG CTCACGGATA 3960

GGCGCATCCG GTGCCACAAT GGCATCTCCA GAACCAGAAA CCGTACAAAA AGTACAGCCC 4020

CCATGAGCCA CAGTCCCATC ACGATTGGGA CAATCAAATC CCGCATCAAT AGGGACTTTA 4080

AAAGTCTTTT CTCCAAAGAG TTTTCGATAA TAATCATTCA AGGTATTATA AGATTTCATG 4140

ACTTTCATTA TAACAAAAAT CACCCACAAT CTCAAAAGCC TGACTTTCCT ATAAATTCCT 4200

CTGTTTCTCG TTTCCATTAG CCTTTTTTTA TGATACAATA TGGGTATGAT TTTAATGAAA 4260

TTAGCATCTA TTTTATTATT GATACTGACC TTAGTCGTCT GCATTATCCT AACCAAACTT 4320

TTTAGATTAA AAAAACTAGG ACGAAACTTT GCGGATTTGG CTTTTCCAGT CTTGGTATTT 4380

GAGTATTACT TGATTACAGC TAAAACCTTT ACCCATAATT TCCTCCCTAG ACTGGGGCTA 4440

GCCCTCTCGA TCCTAGCCAT TATTCTCGTC TTTTTCTTCC TTTTGAAAAA ACGCAGCTTT 4500

TACTACCCTA AATTTATCAA ATTCTTCTGG CGTGCAGGAT TCTTATTAAC CCTTATCATG 4560

TATATAGAAA TGATTGTTGA ATTGTTCTTA ATGAAATAGT CGAATCCCTA AGCATTTTCT 4620

AGGGATTTTT GCTTTCTCTA CAAAATAGTA TAGACAATAA CACTATACAA TTTTATACAA 4680

AGAAAAGAGT CTGGGACAAT AGTCTCTTAT ATCCAAAAAG GCAACGGATT TGCCGTTGCT 4740

TTTTTGGATG GTTACGATAG TCTTGGTAAA ATAGAATTGC CCAATAAACC ATTTAGAAAG 4800

GCTATCCCAT GCATATTCAC TATAACACAA ATCAAACAAC TTTACCACTA GAAATCAGTT 4860

CCTTCTTACC ACAAGATCAT CTCGTTTTTA CTATTGAAAA AGTGGTGAAT ACCTTGGAGG 4920

AACGTCACTT CTACACCTCC TATCATGCCT TTGATCGCCC GTCTTATCAC CCTAAAATGC 4980

TTGTATCTAC TCTTCTATTT GCCTATTCAC AAGGGATTTT CTCTGGTCGA AAAATTGAAA 5040

AATGGAAGAG TTAGTGACCT TAGATTGTTT GTTTATTGAC AGAACTAAGA TTGAAGCCAA 5100

TGCCAACAAG TATAGTTTTG TGTGGAAGAA AACGACAGAG AAATTCTCCG CCAAACTTCA 5160

AGAACAGATA CAGGTCTATT TTCAAGAAGA AATCACTCCC CTTCTGATTA AATATGCCAT 5220

GTTTGATAAG AAACAAAAGA GAGGGTATAA AGAGTCAGCT AAAAACTTAG CGAATTGGCA 5280

TGGCAATATC CGTTGCGCGC ACACTTAGCA AGAAGGTATG GAGGAAGCGG TTGGTCATGG 1800

CTGTTAAAAA TCCAATTAAA AAACCGCCTA GGATTGAAAA GATGAGCGAA CTGCTAGCTT 1860

GCCCCAGAGA AAAAGCTCCA GTTGTCCAAG CTGTCAAAGC TACCTGAAAA GCCACCAAAC 1920

CAGAAGCATC ATTCAAGAGT CCTTCGCCCT TAAGAATATT GGACACGCGC TTAGGAAAGC 1980

TAAAACGCTC CGAAAGAGAG GCAAAGGCCA CCAAGTCCGT AGGACCAAGG GCTGCCCCAA 2040

CAGCCAAGCA AGCTGCCAAG GGAAGGCTGA ACCAAAGAAG ATGGGCCAAG CCACCCAAAC 2100

TCAGGGTCGA GATAAAAATC ACTGGAAATA TGAGATAAAC AATGATTCGC CAGTGTTTTA 2160

AAATAGCCGT AACATCTGCT TCTTCAGCCT CTCGGAAAAG CAAGGGTCCG ATAACCAGTG 2220

CCAAAAACAA CTCCGTATTA AGGTGAAAGT CAGTATTGGG TAAAAAGAGA CCAATCACAA 2280

TTCCCAAAAG AATTTGCACC AAAGGGAGAG GCAAAAAGGG CAGGAGCTTA TTGGTTGTAC 2340

TTGAGACAAT CAAAACCAGT AAAAATAGGA TGAGGTAAAT CAGTAATTCC ACGCACGTCC 2400

TCCTTAATCT TTTTTACAAC AGGATTCAAA TATCTCCTTC TGCTCTTTGA TTTTTTGGTC 2460

AATCTTGGAA CAGTCTTTGT GCTCAATTTT TCTCTGGCAC CGTTCCATTT CAAGAGCAAC 2520

TAATTTTTTC TTGATTTTAA GCATTTTTTT GCTCATATGC GCTTGGTCTA GCACGCCCAT 2580

CGCTCGTTCG TGGTGGGTTG ATTCAACAAA ATTCTGGCGC ATGGCATCCA GCTTTTCGTG 2640

TAAGTATTGT TTATCCATGT CTGTATCTCT CTAATTTTTC AATCATCACT AAAAACGGCG 2700

GGTTGTTGAC TTGGTTTAAA GTTCGGTAAA TGGCAGCTGT GTACTCTTGT TGGTTCAACT 2760

GGATCACAAA ATCCAAGACA GCATCTCTCT CGAGATCGCC TCCTTCATGA CCATAGTAAA 2820

TCATAATAGC AATTCGTCCA CCTTTGACAA GTAAGCCACA TAGCTTTTCT AATGCCTCAA 2880

TCGTTGTCTG CGGTCGGGTG ATGACAGACT TATCAGCTGC CGGCAAATAG CCCAGATTAA 2940

AAATCCCTGC CTTAGCTTTT ATCACAAACT GGTCCAGTGT CTCATGGCCT TGCAAGATTA 3000

ACTGGGCATT TGTCAAGTCA GCCTGATGCA AACGCTCTTG GGTCTTTTCC AAGGCTTGCT 3060

TCTGAATATC AAAGGCATAG ACTTGCTTGG CTAGCTTGGC TAAAAAAAGC GTGTCATGAC 3120

CATTTCCCAT AGTCGCATCC ACTACGACAT CCTCTTTTGT CACGACCTCA GCCAAAAAAT 3180

CATGTGCCAT CTCAAGTGGT CTTTTCATTT TCAAACTCCT GTTTTACAGC CTTGCATCCT 3240

TGAACACTTC CACGACGTCG CATCTCCATC TCAATGCTGT TGAGGACTTC CCATTTATTG 3300

AGGCTCCACA TAGGACCAAG CAGCATATCC CTAGGCGCAT CTCCTGTAAT TCGATGGATG 3360

ACGATATGTT TGGGAATAAT TTCCAGTTGG TCACAGATGA CCCTGACATA TTCGTCCTGA 3420

CTCATCAATT GTAAACGCCC CTCATGGTAA TCTCGTTGCA TACGAGTATT TGTCATAAGA 3480

TGGAGCAAAT GCAGTTTAAT CCCTTGAATA TCGTTATCCG TGACACAACG GCGGACATTT 3540

CCGGAGGATA TTGACCACCA CCAAAAGCAG GGGGAAAATC GAAATCAACC AATAGTAGGC 60

TACTGCGACA CTGGTCAACT CACTATCTGA TGCTTGATAA TAATGCAAAA AAGCTTTTAA 120

TAAAGGTTTG TCTATCAGCT CTTTCCACCA CTTTTTCATG TCATACTCCT TCACTTATAA 180

TCTTATACTC AATGAAAATC AAAGAGCAAA CTAGAAAGCT AGCCGCAAGC TGCTCAAAAC 240

ACTGTTTTGA GGTTGTAGAT AAGACTGACG AAGTCGATCA CATACATACG GTAAGGCGAC 300

GCTGACGTGG TTTGAAGAGA TTTTCGAAGA GTATTAACTA ATTTCTTCTT ACCAATTCCA 360

CCATATCATA CGGTAGGGTA TTGGCAGCTT CCTTCAAGGA ATAGTTCTCT AAGTTATTTA 420

CATTTTGTCG TAATTTCTTG GCATACTTAG TCGTAATCAA TCGTTTTTCT TCGTATTCGA 480

AAATCAACTT GCGCTCCAGA TAATAGCCTC TCAGCATTTC ATCGATATTG TTGGGTTTGA 540

CACGATTGAT AACCCGTTCG ACAAAGGCAC CACTGCTGAT AATAGCTGTT TCTCGAAGAC 600

GAGACTCCTG CATAAAACTA ATCAAAGAGC GTCTGTAGAC TCCCTTCAGG TTTTCCAAAC 660

TTTCAATAAT CATCTCTGTA TTGGCAAGAT AGAGCTCTGC AATTTGGTCA TAATCAAGAG 720

CACGGAGACG GCTTTGCTCC TTGTTCTTCC AGCTACGGAA GGTCTTTCCG AGAGTAAAAA 780

CTTCATGAAG GAGAAAACGT AAAATCCTCA AGGAAACAAG AAAATAATAG GTCAGTCTTG 840

AGGCAAGTTT ACGATTGATT CCTTGTTCTA TATTTTTCAG ATAACGTTGG TAAACTCGGT 900

AAGCACGATT GCTAATGTTC CCCTCTTCAT AGGCCTGTTC CAAACCATCA CTTTCAATAC 960

TAAGAATCAA GAGTTTCAAA GCAGCCCAGT CTTCTTGATC ATCCTGGTTT TCTTGGCTTA 1020

AAATGAGATT TTCAATACGT CCATGATAAT TGTCAATAGC CGCATAGAGG GGAAGTTTAT 1080

TTCTGGTGTC TTCCAACTCT TTTTCCAACT CTAGCGTTAC TTCATTCAAA ATGGCGATAT 1140

GCATAAGATA ATCCTTGCTT TCTTCCTCTT CATCAGAAAG ATGAGGCAAG ACCAAGAGAC 1200

CTGTTAAAAA GCTAACAAGC GTCACACCTG CAACAAGGAA AAGCAAAAGA GGATACTCCT 1260

GTTCTAGATT ACTTGGTATC AAGAGAATCG TAGCAATCGA CACCGTTCCC TTAACACCTG 1320

AAAAGGTCAA GAGAAACATG TCCTTCATAT ACTTATTTAG CTTTTTCTTG AGGCGTCGGG 1380

TTCTATAGGC ATAATAGCCA TAGATCATAA TAAAACGAAT GACAAAAAGG ACAAAGGTAA 1440

GGGCGATAAG AGATAGCAAT AAAAGTAGAG GATTATAGAT TGGATTGGTC AAGATAGGTT 1500

CTGCTATCAT TTCCAACTCC ATCCCTAAAA TCACAAAGAC AGAACCGTTG AGCATAAAGG 1560

TCACTGTATG CCAGACCGTC TCGGTCACCG TATCCACTTG GGCTTCGAGG AGCGTGATTT 1620

TCTTGAAGCG ACTTGCCTTT AAAATTCCAG CAACTACGAC GGCAATAATA CCTGAAACAT 1680

GAACTTCTTC TGCCAGAAAG AAGGTCACTA GAGGCAAACT CAATTCTAAT AAAAGTTCAC 1740

CTTCTCTTTC AAGCGTAATA CCTGAGTGTT CCTTGACTTT TTCGATAACC GATTGGATCA 1740

AGTCCTCGTA GTCTTTGGCC GTTCCATCTG CGACATTGAT CATAAATCCT GCATGCTTTT 1800

CTGACACTTC TACGCCACCG ATACGATAGC CTTTCAAGCC AGCTTCTGAA ATTAACTGAC 1860

CTGCAAAATG CCCGACTGGA CGCTTAAAGA CCGAGCCACA AGATGGGTAT TCCAAAGGTT 1920

GCTTGAGTTC ACGTAGGTGC GTCAAGCGGT CCATTTCCTG CTTGATAACC TGATGGGTTC 1980

CTGGAGCTAG GGCAAATTTA ACTGACAAGA CAACTGCACC AGACTCCTGA ATAGCTGAAT 2040

GACGGTAACC AAAAGCCAAG TCTTTAGCAG ACAGGGTTTC GATTTCTCCA TCCTTGGTCA 2100

AGACCTTACA AGACTGCAAG ATGTGAGCAA TCTCGCCACC ATAGGCACCC GCATTCATAA 2160

AGACAGCACC GCCAACGCTT CCTGGAATAC CACAAGCAAA CTCAAAGCCA GTTAAACTAT 2220

GACGGAGGGC AATGCGAGTT GTTTCAATCA AGTTAGCCCC AGCTTCTGCT TCAATGGTAT 2280

AGCCATCAAC AGAAACGTTA TTGAGCTTGT CACACAAGAT GACAAATCCA CGAATCCCAC 2340

CATCACGAAC GATGATATTG CTTGCATTGC CAAGAACCAT CCAAGGGATA TTTTCTTGGT 2400

TGGCAAATTT CACAACGCGA GCCAACTCAA AACGATTTCG TGGAAAGACC AAATAATCAG 2460

CCTCTCCACC TACTTTTGTA TAACTATAGC TATGCAAGGG TTCCTTAAAA CGGATATCAA 2520

TTCCTTCTAA GATTTCAAGC ATTTTTTCTC TTACAGACAT GTCACTCTTC CTTTTACAAA 2580

ATTCATTCCA TTATACCATT TTTAGAGACA TTTGACGACC ATAAAAATAC CTTGTTTGGA 2640

TTTTGCATAA GAAAAAGAGG TTCCCCCCTT TTTATGATTT TTTACAAAAG ATTTCCTTGG 2700

TTCCATAGGC GACCAGAACG AGCTCCAGTG CTAGAATCAC TTCAACCAAG ACTGGATTTG 2760

TCAACCAGCC TACTTGGAAA AGAGATGGTG CCAGATCAAA GAAGGCATGC AAGCCATAGG 2820

CTGCTAGGAG ATAAATCCAT TTCTTCTGGC GAACAGCTTG GTAAACCCAA ACTGTCAAAA 2880

GTAATTGGAA ACCAAGCGCC AAGATTCGCT CAAAACCAAG CAAATAAATC TGCCAGACCG 2940

AAAGTGACTG AATGGTTTTT AACATATTTT CAGACAGTAA TTGCATAACC TGTGGATTCT 3000

GAGTTTGAAC TGCCGAAAGA ACAATGTAAA GATTGAGTAA ACTAGTAAGG CCTAGAAAAA 3060

TCAACTCCAA GCCACCATGC CCC 3083

(2) INFORMATION FOR SEQ ID NO: 139:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15363 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

AGCAACTGTT GTGAACCAAT TCCGATAAAT TCCAAGAATT GGTTAATAGA GCCATTTTGA 60

CCAAAAATCC CGATAAAAGC ATAGGCTTTA AGGAGCAAAT TGATCCAGGT AGGAAGGATA 120

ATCAGCATGA GCCAGAGTTG ACGGTGTTTG AGACGGGTCA AAAAGAGGGC CGTCGGATAA 180

CTGATAAGCA GTGCCACAAA GGTCACAATG CCTGCATAAA GCACTGAGTT GAAACTCATT 240

TTAAGATAGG TCAAGTTTTG TGACGCAAAG TAAGATTTGT AATTTTCTAA ACTGAACTGG 300

CCTTCGATGT TGAAAAAGGA TTGACCGAAA ATCAAGACCA AGGGTGCCAA TACAAAGAGC 360

GCAATCCAAA GCATGTAGGG TACTACAAAG AGTTTAGAGC TTGTTTTCTT CATCTCTTTC 420

CTCCTCGATT GCATTGATCA AACCTGCTTC TTGCTCTTCG ATTTCTACGT ACTCCTCAAT 480

ACGAGCATCG AACTCTTCTT CGGTTTCATT GAGACGCATG ATGTGGATGT CTTCTGGTTC 540

AAAGTCCAGA CCGATTTCCT CACCCACGAT AGCCTTACGG GTTGAGTGGA TCATCCATTC 600

ATTTCCAAGT TCGTCATAGG CGATAATTTC ATAATGAACT CCACGGAAAA GCTGGGTATC 660

GACCTTAACT TGGAGCTTGC CTTCTTCAGG AAGGGTAATG CGCAAGTCCT CTGGACGAAT 720

AACGACCTCA ACAGGTTCAT TTGGCTTCAT CCCACCATCA ACCGCTTCAA AGCGTTTGCC 780

GTTAAATTCG ACCAAGTAGT CCTCAATCAT GGTACCTGGC AAGATGTTTG ACTCCCCGAT 840

AAAGGTGGCA ACAAAGTGGT TGATTGGCTC ATCGTAGATG TCCACAGGGG TTCCAGACTG 900

GACAATCTCG CCATCATTCA TAACGAAAAT CCAGTCACTC ATGGCAAGAG CTTCTTCCTG 960

ATCGTGAGTG ACAAAGACAA AGGTAATGCC CAATCGTTGT TGTAATTCAC GCAATTCGTA 1020

CTGCATGTCT GTTCTCAATT TCAAGTCCAG CGCTGATAAA GGCTCGTCCA ACAAGACCAC 1080

ACGGGGTTGG TTGATGATAG CACGGGCGAT GGCCACACGC TGACGTTGTC CTCCAGAAAG 1140

TTTGCGGATG GAACGTTTTT CATAACCTTC CAACTGAACC ATCTTGAGAA CTTCCGCTAC 1200

ACGCTGCTCG ATTTCTTTCT TATCAATTTT ACGCAAGCGA AGTGGAAAGG CAACATTTTC 1260

AAACACATTC ATATGTGGGA ACAAGGCATA GGATTGGAAG ACGGTATGTA CGTCGCGCTT 1320

GTTGGTTGGA ATATCATTGA TACGAACACC GTCTAGCATG ATATCTCCTG TCGTCGCATC 1380

CAGTAAACCT GCAATAATGT TTAGGATAGT TGATTTCCCC GAACCAGATG CACCTAGAAG 1440

GGTGTAGAAT TTCCCTTCTT CCAACTCAAA GTTGATGTCT TTGAGAACCT TGGTGTTGCT 1500

GTCTTCAAAA ACTTTAGAGA CGTTTTTGAA TTCGATAATT GGCTTTTTCA ATTGGCATAA 1560

ATTCCTTCTT TTTCATAGAT TAACCGATCG GGGCTCTGTC AGGTCCCCAC TACCTCTTGC 1620

AGGGAGTAAA ACCACCTGCA TACATCTTCG CTACCGATAG GCTTTCACCC AAGATCCGGA 1680

ATTAATTTCT AGAAATATGT TAGAAATTGG TTTGAATTCC GCAATCAATT TGTTCAGTTT 11280

TTATTTCATT TCATTTTATT TAATTAGATT TTCCAATTTT TTAATTCAAG CTAAAAATCC 11340

CCAATCGTAG TGATTGAGGA TTGAGTAAAT AAATCTTAAA CAATACCTTG TGCAATCATG 11400

GCATTTGCTA CATTTTCAAA GGCAGCAATG TTAGCTCCTG CAAGGTAGTC TTTATCAAGA 11460

CCGTATGTTT CTGAAGTCGT TTTAGCTGTG TTGAAGATGT TTGTCATGAT GTCTTTGAGA 11520

CGGCCATCAA CTTCTTCACG AGTCCATGAG AGGCGAAGAC TGTTTTGGCT CATTTCAAGA 11580

GCTGAAACGG CTACACCACC AGCGTTGGCA GCTTTTGCAG GTCCGTAGAA GATACCATTT 11640

TCTTTGTAAA CTTTGATGGC ATCAAGGTCG CTCGGCATGT TGGCACCTTC AGATACACAG 11700

ATAACGCCTT GAGCAACCAA ACGTTTAGCT GCTTCACCGT TGATTTCGTT TTGAGTGGCA 11760

CATGGAAGAG CAATGTCATA GTTTCCAGCG TAAGTCCATA CAGTACCTTC GTGGTAGGTT 11820

GCAGTTGCTT TTTCAGCTGC ATACTCAGTC AAACGAGCAC GACGTTTTTC TTTAACATCA 11880

ACCAAAAGAT CGAAGTCGAT ACCATTTTCA TCGATGACAT AACCATTTGA GTCAGAAACA 11940

GAAATAACAG TTGCACCGAG TTCAGTTGCT TTTTGAAGAG CATATTGAGC AACGTTACCA 12000

GAACCTGAAA TAACGACTTT CTTACCAGCA AAGCTGTTAC CGTTAGCTTT GAGCATTTCT 12060

TCAGTATAGT AAACCAAACC GTAACCAGTT GCTTCTGGAC GAATCAAGCT ACCACCAAAT 12120

CCAAGAGGTT TACCAGTCAA GACACCAGCA TCAAATTGGT TAAGACGTTT GTATTGACCG 12180

TAAAGGTAAC CAATTTCACG TCCACCAACA CCGATATCAC CAGCAGGTAC GTCAAGTGAT 12240

GGTCCGATGT GTTTTTGCAA TTCAGTCATG AAGCTTTGGC AGAAGCGCAT CACTTCAGCA 12300

TCTGTTTTAC CTTTAGGATC GAAGTCTGAT CCACCTTTAC CTCCACCGAT AGGAAGTCCA 12360

GTCAAGACAT TTTTAAAGAT TTGTTCAAAT CCGAGGAATT TCAAGATCCC TTGGTTTACA 12420

GTTGGGTGGA AACGAAGTCC ACCTTTGTAT GGTCCAACAG CTGAGTTGAA TTGAACACGG 12480

TAACCACGGT TTACTTGAAT TTTTCCATCA CGGTCAACCC AAGGAACACG GAAAGAAACC 12540

ACGCGCTCAG GCTCAGTAAT ACGTGCCAAG ATATTTTCTT CGATATACTC AGGGTGTTTT 12600

TCAAATACAG GTTCTAAAGT GTTGAAAAAT TCTTCAACAG CTTGGAGGAA TTCAGCCTCG 12660

TGCCGG 12666

(2) INFORMATION FOR SEQ ID NO: 138:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3083 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

TATCGAACTT ATGAACAAAT TGCGGCTGTT TTTGGTATTC ACGAAAGCAA CTTAATCCGT 9540

CGGAGCCAAT GGGTTGAAGT AACTCTTGTT CAAAGTGGTG TTACGATTTC AAGAACTCCT 9600

CTCAGTTCTG AGGACACGGT AATGATTGAT GCGACGGAAG TAAAAATCAA TCGCCCTAAA 9660

AAAAGAATTA GCGAATTATT CTGGTAAAAA GAAATTTCAC GCTATGAAGG CTCAAGCGAT 9720

TGTCACAAGT CAAGGGAGAA TTGTTTCTTT GGATATCACT GTGAACTATT GTCATGATAT 9780

GAAGTTGTTC AAAATGAGTC GCAGAAATAT CAGACAAGCT GGTAAAATCT TGGCTGACAG 9840

TGGTTATCAA GGGCTCATGA AGATATATCC TCAAGCACAA ACTTCACGTA AATCCAGCAA 9900

ACTCAAACCG CTAACAATTG AAGATAAAGT CTATAACCAT GCGCTATCTA AGGAGAGAAG 9960

CAAGGTTGAG AACATCTTTG CCAAAGTAAA AACGTTTAAA ATGATTTCAA CAACCTATCG 10020

AAATCATCTA AACGCTTCGG ATTACGAATG AATTTGATTG CTGGTATTAT CAATCATGAA 10080

CTAGGATTCT AGTTTTGCAG GAAGTCTATT ATCAAAAATA CCATCAAGAT TATATAAGAT 10140

TGATACAGGA AAAGTTTTAT TTGATGGTGT AAATATTAAT CAAATAGATA AAAAAATATT 10200

AAGTCAAAAT TTAGGAGTAG TTCCACAGGA TTCATTTTTA TTGAACCGAA GTATTCTTGA 10260

TAATATAACT TTAAAGCACG AAGTTACTTC ACAAAAGATA GAGGAAGTTT GTAAAGCAGT 10320

TCAAATCTAT GATGAAATCA TGGCTATGCC GATGAAATTT AATACTATCA TCTCAGAGAT 10380

GGGGTCAAAT ATTTCAGGTG GGCAAAGGCA ACGGATAGCA CTGGCACGTG CATTAATAAA 10440

TAATCCTAGT ATTGTAATTT TAGATGAAGC AACTAGTGCA TTAGACACTA TTAATGAGGA 10500

AAGAATAACA AAGTATATAC AAAGTCAGGG CTGTACTCAA ATAATTGTAG CTCATAGATT 10560

GTCAACGATT AAGGATGCGG ATGTTATTTT TGTAATGAAA GGTGGTAAGA TTGTTGAGTC 10620

AGGAAATCAT AAGTACTTAA TGGATCTTGG TGGAGAGTAC TACAGCTTAT ATACAAAAAG 10680

GAAATGAGGT GTAAAGAAAA TGAAGAAAGA AAATGAATAT GTAATTTTAA CAACAGCCTC 10740

ACTAGGGGTG ATGATTGGAA TAGTGTTTGC AATTTTTTTA GATTTTCCAG TTGAATATGG 10800

TATTTCTTTA GGCTTGTTGA ATGGAATAGT ATTGGGTTCG CTGATTGTTT ACAAAAACAA 10860

TAAGAATTAA GCATAATTTT TTGCTGTAAA CTAAGGAGTA GAGATGGCTA TAGTTGAAAT 10920

TATAAATCTA ACAAAAAGCT TTAAAGATAT TGAAGTTATT CATAACACTT AAATAATAGA 10980

GCAACTACAG TAGTAGCTTA AAAACATGAT TAAATCGCTA TTCTTAGGAG TAGCGGTTTT 11040

TCTTTTTGTT TAATACTCTT TGAAAATCTC TTCAAACCAC GTCAGCTTTG CTTTACCGTA 11100

CTCAAGTACA GCCTGCGGCT CGCTTCCTAG TTTGCTCTTT GATTTTCATT GAGTATAAAA 11160

AGGGTCAAGT AAGTATAGTA AATTGAAATA AGATATGAAC AAATCGATTA GAAAAGTCAA 11220

AAATGAGGAC TTTCTTTTTA TACTCATCTG CTTTCAAAAA GCATTCTAGT CCATCTCCGA 7740

TTAACGATGG ACTTTATCAC CTCCTTCTCC AGTCCTTGTA TAACATCTTG GAGTTGATTC 7800

ATGACATCTT CCAAAGTTTA AAAGGCTTTA TTCTTAAATC CACGTTTACG AATCTCTTTC 7860

CACACTTGTT CAATGGGGTT CATCTCTGGT GTGTATGGAG GAATAAATGC AAAGCCAATA 7920

TTAGTCGGAA TCTTTAAGGT ACTTGATTTA TGCCATATAG CATTGTCCAT AACGAGTAAA 7980

AGATAATCAT CTGGATAAGC TTGTGAAATC TCCTATTCCT AAAGCCCCTT TAGCGCATAA 8040

CTTTGGCTCA GCTTCTATTA TCGCTCACAC CATCCATCAG AAGTTTAATC TGAAGGTACC 8100

CAATTATCGC CAAGAAGAAG ATTGGGCTAG GATGGGTTTA CCAATCACAC GTAAGGAAAT 8160

CTCTAATTGG CATATCAAGG CGAGTCAATA CTATTTGGAG CCCCTTTATA ACCTCTTGCG 8220

AGAGAGACTA TTGACTCAGC CCTTACTTCA TGCGGATGAA ACTTCTTATA GGGTGCTAGA 8280

GAGTGATAGT CAGCTGACTT ACTATTGGAC TTTTTTGTCA GGTAAAGCAG AGAAACAAGG 8340

GATTACGCTT TACCACCATG ATCAGTGTCG AAGTGGTTCA GTAGTACAAG AATTCCTAGG 8400

AGATTATTCT GGCTATGTGC ATTGTGATAT TTTGCGGCAG TAACTTAGGA CTTTAGTCCT 8460

CTAGTTCTGC CTATGCGATA GCAGTCCAAG GTTTAGGAGC AAGGCGACGC TAAGCTTGGT 8520

AAACTTCGAA CCGCTCGTCT GCTTATCGTC AACTGGAAGA AGCTGAACTT GTTGGATGTT 8580

GGGCGCATGT GAGAAGGAAG TTTTTTGAAG CGCCCCCCCA AGCAAGCGGA TAAATCATCC 8640

TTAGGAGCTA AAGGTTTAGC TTATTGTGAT CAGTTATTTT CCTTGGAAAG AGACTGGGAG 8700

GCTTTGCCAG CTGATGAACG ACTACAGAAA CGTCAAGAAC ATCTCCAGCC CTTAATGGAA 8760

GACTTCTTTG CTTAGTGCCG GCGTCAGTCA GTTTTAGCAG GTTCAAAACT AGGAAGGGCA 8820

ATTGAATACA GCCTCAAGTA TGAAGAAACC TTTAAGACCA TTTTGAAAGA CGGACATCTG 8880

GTCCTTTCCA ATAATCTAGC TGAACGCGCC ATTAAATCAT TGGTTATGGG ACGGAGTAAA 8940

AGAGTCCAGT GGACTCTTTT AGCCTAAGCT CAGTTTAAAA AAGCGAGGGT GGTTATTTTC 9000

TCAAAGTTTT GAAGGAGCTA AAGCAAGAGC TATTATTATG AGTTTGTTGG AAACAGCTAA 9060

ACGTCATCAA TTAAATAGCG AGAAATATCT ATCCTATCTT CTAGAATGTC TTCCAAACGA 9120

GGAAACTCTC GTAAACAAAG AGGTTTTAGA GGCTTATTTA CCATGGACTA AAGTTGTACA 9180

AGAAAAGTGC AAATAAGAAA TCTCCAGATT AGGAACTATC CGTGAGTTCT CCAGTCTGGA 9240

GATTTTTCAA TAGACTTCCT GCGAAACAAA ATATGGTATA ATAGTTCTAT GAATGATGAA 9300

GCAAGTAAAC AACTAACCGA TGCACGATTT AAGCGTCTTG TTGGTGTTCA ACGCACGACT 9360

TTTGAAGAGA TGTTAGCTGT ATTAAAAACA GCTTATCAAC TTAAACACGC AAAAGGTGGA 9420

CGAAAACCTA AATTAAGTCT AGAAGACCTT CTTATGGCCA CTCTTCAATA TGTGCGAGAA 9480

GGAGAATCGG CTTTATCTGA TGTAGATAAA TTGTATGTGA GGTTTACAAA GCGTTTTGAA 6000

GAAGAGTACA TAAACCAAGG ATTTTATAAA AATCGAAATA TAGAAGATAC GTTGAATCTT 6060

GGGTGGGAAT TACTATCAAT TCTTCCTAGA ACAGAGTTAA AACGTATCAA AGATGATTTG 6120

CTTGATAAAT ACTTACCTTT GGTAGAAGTT TAATCCGGAA ATGGAGTGAT TATCTATGGT 6180

ACGTTTGAAT GTAAAACCAA CTCGTATGGA ATTGAATAAC TTAAAGGAAC GTTTGACAAC 6240

AGCTGAACGT GGACATAAGT TATTAAAGGA TAAAAGAGAT GAATTGATGA GGCGATTTAT 6300

TTCTTTGATT CGTGAGAATA ATCAACTTCG GAAAGAAGTG GAAAGTTATC TAATTGATAA 6360

TCTAAAATCC TTTGCAGTTG CTAAATCATT AAAGAATTCT CAAATGGTGG AGGAATTATT 6420

TTCAATTCCA TCGAAAGAAA TTGAATTATT TGTTGAGAAA GAAAATATCA TGAGTGTAAC 6480

AGTTCCTAGA ATGCATATGA ATATTACTTC TCAAAATGAG AACAGTGAAT ACAGCTATTT 6540

ATCTTCTAAT AGTGAAATGG ATGATGTATT TGCTACAATG AATAGTTTAA TTTATAAATT 6600

ACTAAGACTG GCAGAAGTTG AAAAAACGTG TCAGTTAATG GCTGATGAAA TAGAAAAAAC 6660

ACGTAGACGT GTAAATGGTT TAGAATACTC GATTATTCCA AACTTGTCGG AAACTATTCA 6720

TTATATAGAA TTGAAACTAG AGGAGGCAGA AAGAGCCAAT TTAGTTCGTA TTATGAAAGT 6780

GAAGTAGATC CTTTATTTAG ATTATTAATT AGATGAACAA ATATCAGCTT GGATAAGGCT 6840

TTAAGCCTTT CTAAGCTTTT TTTATTGACA GTATCAGGAT ATCTTTTTCA AAATTTTGGT 6900

TTGTTAGATA ATGAAAATGT TTCTACTAAT CTAGATTTAG GATTAGTAAA TCGTAAATGT 6960

AATTATATAG AAAGTAAGCG CGTCATAACA AGGTATCTAT CATTCATGGA GCTCCTCCTG 7020

TATACTATTA GTAAAGTAAA ACTATTGGAG GATATTTTAA TGCCACAACC TATTGTTCCT 7080

GTAGAGATTC CACAATCTCG TCGTTTTGAT TCTAAAAAGA GAAATGATAT TCTGCTTAAA 7140

ATTCGTATTG GCAAGCTTGA AGTAAGTTTT TTTCAATCTC TCAATCTCGA AATGGTAGAA 7200

CAGCTTTTGG ATAAGGTGTT GCTCTATGAC AATTCATCTA TCTAGCCTAG GGGAGGTCTA 7260

TCTCGTGTGT GGGAAAACTG ATATGAGACA AGGAATCGAT TCACTGGCTT ATCTGGTTAA 7320

AACCCACTTT GAATTGGATC CTTTCTCCGG TCAAGTCTTT CTCTTTTGTG GTGGACGTAA 7380

AGACCGCTTT AAAGTCCTTT ACTGGGATGG TCAAGGATTT TGGCTACTAT ATAAACGCTT 7440

TGAGAACGGC AGATTGATTT GGCTAAGTAC AGAAAAGGAT GTCAAAGCTC TCACACCAGA 7500

ACAAGTAGAC TGGCTTATGA AGGGCTTTTC TATCACTCCA AAAATATAGT AGATTGAAAC 7560

TAGAATAGTA CACCTCTGCT TCTAAAACAT TGTTAGAAAT CGATTTTACT GTCCTGATCG 7620

ATTTGTCCTG TTCTTATTTC ATTTTACTAT AAATCCATCA GAAAGTCGTG ATTTCTATTG 7680

ACGATTACTG CTATTGGAGC TGTATCGCCA CCTGGTGGAG ATATTTCAGA ACCAGTTACT 4200

CAAAACACTT TACGGATTGT GAAAGTTTTT TGGGGGCTTG ATGCTCCGTT GGCACAGCGA 4260

CGTCATTTTC CTGCAATTAA CTGGCTTACA TCTTATTCAC TATATAAAGA CAGTGTGGGC 4320

ACTTATATAG ATGGTAAAGA GAAGACAGAT TGGAATAGTA AAATAACTCG TGCGATGAAC 4380

TACTTACAAC GGGAATCTAG TTTAGAGGAA ATTGTTCGTC TTGTTGGAAT TGATTCTCTG 4440

TCTGATAATG AACGACTAAC GATGGAAATT GCTAAACAAA TTCGAGAAGA TTATTTGCAA 4500

CAGAACGCTT TTGATTCGGT AGATACATTC ACTTCGTTTG CAAAACAAGA AGCAATGCTA 4560

AGTAATATTC TCACTTTTGC TGATCAGGCA AATCATGCTT TAGAGTTGGG TTCTTACTTT 4620

ACAGAGATTA TGGAAGGTAC CGTGGCAGTT CGAGACCGTA TGGCGAGAAG TAAATATGTT 4680

TCAGAAGATA GATTAGATGA AATCAAAATT ATATCAAATG AGATTACACA TCAAATTCAT 4740

TTGATATTAG AAACAGGAGG TCTATAAATG AGTGTTATAA AAGAATACAG AACTGCTAGT 4800

GAAGTTGTTG GGCCTCTTAT GATTGTTGAA CAAGTAAATA ATGTGTCTTA CAATGAGTTA 4860

GTTGAAATTC AACTTCATAA TGGAGAAATT CGTCGTGGAC AAGTTTTAGA GATCCACGAA 4920

GATAAAGCAA TGGTTCAGCT TTTTGAAGGA TCTAGTGGAA TAAATTTAGA AAAGTCTAAA 4980

ATTCGTTTTG CTGGTCATGC ATTAGAATTG GCTGTATCTG AGGATATGGT TGGTCGTATT 5040

TTTAATGGGA TGGGAAAACC AATTGATGGT GGACCAGATT TAATTCCAGA GAAATATTTA 5100

GATATTGATG GTCAAGCTAT TAATCCTGTA TCTAGAGATT ATCCAGATGA ATTTATTCAG 5160

ACAGGGATCT CCTCTATTGA TCATTTGAAT ACTCTTGTAC GTGGTCAAAA ATTACCAGTA 5220

TTTTCAGGTT CGGGCTTACC TCATAATGAA TTAGCTGCTC AGATAGCAAG ACAAGCGACT 5280

GTTTTAAATT CTGATGAAAA TTTTGCGGTT GTATTTGCAG CAATGGGTAT TACTTTTGAA 5340

GAAGCTGAGT TTTTTATGGA AGAACTCAGA AAAACAGGAG CGATCGATCG TTCGGTTTTA 5400

TTTATGAACT TGGCAAATGA TCCTGCAATT GAGCGTATTG CAACTCCCCG CATTGCTTTA 5460

ACTGCGGCAG AGTATCTAGC TTTTGAAAAA GATATGCACG TTCTAGTTAT CATGACGGAT 5520

ATGACTAACT ATTGTGAAGC GTTACGTGAA GTCTCGGCAG CTCGCCGTGA AGTTCCAGGG 5580

AGACGAGGCT ATCCGGGATA TTTATATACA AATTTATCAA CTCTATACGA AAGGGCTGGT 5640

CGCTTAGTTG GTAAAAAAGG TTCGGTGACA CAGATTCCTA TTTTAACAAT GCCAGAAGAT 5700

GACATAACAC ATCCAATTCC TGATTTAACT GGATACATTA CTGAAGGGCA AATTATTTTG 5760

TCGCATGAGT TGTATAATCA AGGTTATCGT CCACCAATCA ATGTTTTACC TTCTCTCTCT 5820

CGATTAAAAG ATAAGGGATC TGGAGAAGGT AAAACTCGTG GAGATCATGC TCCAACTATG 5880

AATCAACTGT TTGCAGCCTA TGCCCAAGGG AAAAAGGTTG AAGAGTTAGC AGTAGTATTA 5940

ACAATTTCAT CTTCTGAGTT AGAATTTTTA TGTGATTTAC TATTGTATAA AACTTTAGAT 2460

CAAGGAAGGT ACAATGTAGA GGGGCCGTTA GTTCTTGCTA GATATTTATT GGGATGTGAG 2520

TTTGAAGTAA AGAATCTCAG AATGATCATA TCAGCTCTTC AAAATACAAT TCCCTTTGAA 2580

TCAATAAAAG AAAGGATACG CCCACATTAT GGAAGCTAAT AAGTATAAAA TTGGCATAAT 2640

TGGTAGCCGT GATATTATTT TACCATTTAG CATGATTGGG TTTGATATAT TTCCTGCCTA 2700

CCAAGAACAA GAAGCTATAA ATACACTAAG AAAATTAGCT CAATCTGATT ATGGTGTCAT 2760

TTATATCACT GAAGACATTG CTTCAATGAT ATTAGATACA ATTCGCCATT ATGATTCCCA 2820

AGTTGTGCCT GCTATTATTT TATTACCGAC TCATAAACAA GGTTTAAATT TAGGATTAAA 2880

ACGTATAGAG GATAATGTAG AGAAAGCAGT AGGACACAAT ATTTTATAAT AATGTACAAA 2940

ATTGTCTGTA ATATTATTCT ATAATTTTTG GACTTAGTAA GGAGAATAAC TTTGACTCAA 3000

GGGAAGATTA TAAAAGTATC GGGACCTCTA GTTATTGCAT CAGGTATGCA GGAGGCTAAT 3060

ATTCAAGATA TTTGCCGTGT AGGTAAGCTA GGGTTAATCG GTGAAATTAT TGAAATGAGA 3120

AGAGATCAGG CATCTATCCA AGTCTATGAA GAAACATCTG GTCTTGGTCC GGGAGAACCT 3180

GTTGTTACAA CTGGAGAACC TCTCTCGGTT GAATTAGGGC CAGGATTGAT TTCTCAAATG 3240

TTTGATGGCA TACAACGCCC ATTAGATCGA TTTAAATTGG CTACTCATAA TGATTTTCTA 3300

GTTCGTGGGG TAGAAGTTCC AAGTTTGGAT AGAGATATTA AGTGGCATTT TGATTCCACT 3360

ATAGCAATTG GTCAAAAAGT GAGTACGGGT GATATTCTTG GAACTGTCAA GGAAACCGAG 3420

GTAGTTAATC ATAAAATTAT GGTTCCTTAT GGAGTATCTG GAGAAGTCGT TTCTATTGCA 3480

TCTGGCGATT TTACAATTGA TGAAGTTGTA TATGAAATAA AAAAATTGGA CGGTAGTTTC 3540

TATAAAGGAA CGCTTATGCA AAAATGGCCT GTCCGCAAGG CGCGTCCTGT TTCTAAACGT 3600

TTAATTCCAG AAGAACCATT AATCACAGGT CAACGAGTTA TTGATGCATT CTTTCCAGTA 3660

ACCAAAGGGG GAGCTGCAGC AGTTCCTGGA CCGTTTGGAG CAGGAAAGAC AGTTGTACAA 3720

CACCAAGTAG CTAAATTTGC CAATGTTGAT ATTGTTATTT ATGTCGGTTG TGGAGAACGT 3780

GGAAATGAAA TGACGGATGT ACTGAATGAG TTTCCTGAGT TGATTGACCC TAATACCGGA 3840

CAATCAATTA TGCAACGGAC AGTTCTGATT GCTAATACTT CAAATATGCC TGTTGCTGCT 3900

CGTGAGGCTT CAATTTATAC AGGAATTACC ATGGCTGAGT ATTTTCGTGA TATGGGCTAC 3960

TCTGTCGCCA TTATGGCTGA TTCAACTTCA CGTTGGGCAG AAGCGCTACG TGAAATGTCA 4020

GGACGTCTAG AAGAAATGCC TGGTGATGAG GGTTATCCTG CTTATCTGGG AAGTCGTATC 4080

GCTGAATATT ATGAAAGAGC AGGACGTTCT CAGGTTCTAG GGCTTCCAGA ACGTGAAGGA 4140

AACAGCCTGA AAAGTTTGCC TCAGCTTTGA TATTGCAATT ATTGCCCGGA ACACAAGGAT 660

TATATGGTTT TGTTATTGGA ATTTTAATTT GGTTGCAATT AACTCCAGAA CTTCCTTTAG 720

AAAAAGGCGT TGCTTATTTC TTTGTAGCTC TTCCAATTGC TATTGTAGGA TACTTTTCAG 780

CTAAGCATCA AGGAAATGTA GCAGTAGCGG GAATGCAAAT CTTGGCTAAA AGACCAAAAG 840

AATTCATGAA GGGAGCAATT TTAGCTGCCA TGGTAGAAAC CTATGCAATT CTTGCTTTTG 900

TCGTATCATT CATTTTGACC CTTCGTGTAT AAGAAATAAA TTTGCAATTC AAAGGAGGTG 960

TCTAAATGAG CAATTTAGAA AACTTACGAG AGTCTGTTAT TGAACAAGCT CATGAAAAAG 1020

GGCGTATGAA ATTATTGGAT TCCAAAAAGA AGATTGATGA TGAATTTGAA ATGCAAAAGT 1080

CGCTCATTAT AAAGAAAAAA GAAGCTGAAC ATGAACGAAA GTTAAAAGAA TTGCAACAGA 1140

AATATCAAAT AATTTTTCAA CAATTAAAAA ATAAGGAACG CCAATCAACG TTAGTATCAA 1200

AACAGAAAAT ATTAAAAGAA CTTTTTCAAT CTGCTTTACT AGAAATGGAA TCTTGGAGTG 1260

CAGATAAAGA AATGGAGTTC ATCTATCGAA TTCTGGAACG ATATTCACAA CAAGAGGTCA 1320

TAGTAACCTT TGGGGAACGG ACTTTAGCTA AATTCAATTT GGAACAATTA GAGAAATTGA 1380

AATTCTCTTT TCCAAATTAT TTATTTAGTG AACAACCTAT CTCAAATGAA TCAGGCTTAC 1440

TTATTTCAAT AGGTAAAATT GATGATAACT ATTTGTATAA AACATTAATT GGATCGATTT 1500

CTAAGGAAGA AAGTTCAAGT ATCGCAAATC AAATTTTTAT CAATTAAGGA TGAAATTGGT 1560

TAATCCTTCT TAGAAATTTG GAGTATTCCA ATAAAATTAG AAAGGTATTT TATGGATACT 1620

AATCTTTTTT CAAAAATAAA TACGACGATT TCGGTAAAAG AAAACGATTT TATTACAGAA 1680

GAAAAATTTC AAAAAATTAT ACAATCCAAA GATACGGAGA CATTGGCATT TATCTTAGAA 1740

TCAACTCCCT ATCATTTATC GATTGACATC TTAGAAGATC CTAGTCAGAC AGAGATTTCG 1800

CTAATGACAA AATTAGTCAA TGATTATAGA TGGGCCTATG CTGAAAGTCC GTCTGATATA 1860

ATTGTGACTT TATTTGCTTT ACGATATGTT TATCATAATA TCAAAGTTTT ATTAAAATCT 1920

AAGGCGGCAA TTAAGAAAGA TTTTTCTAAA TTATTAATTC CAATAGGGAT TTTTGATATA 1980

GAAAGTTTAA AACATTTAGT TTCTTCCTTA CATTCAGATA CACTTCCTGA TTTTATGGTT 2040

CGTGAAGTAG AATCAATTTG GAATGAGTAT GAAACTTTTA ATAATATTCG TGTACTTGAT 2100

GTCGGAGCTG ATCTAGCATA TTTTAAACAT CTGAAACTTT TATCTAATGA GTTAGATGAG 2160

GTACTGTCTC AGGTTATTGT CGAAATGATT GACTTTTATA ATATTATTAC TGTAAAACGT 2220

GGTTTATCTC AAAATAAGAG TCATGGGGAT ATTTTACAAT TACTTTCAGA TGAAGGAAGT 2280

ATTTCTGCTA AAGAATTTAT ATACATTGTA GAAAATCAAG AAATATTTGT GTGGTTCAAT 2340

AAAATAAATC CAAGCTTAGA TTCAATCTTT TCAACTTATG AATTGAAGAT GCAGGACGCA 2400

AAACATACTT TGATAATCTT CCTAAAGAAG AACAATTAGC ATGTGAAGTA TTGCAGGCGT 3300

GTTTGGATAC TTCTAGAACT AGAAGGCCTG AATATGCAGA GTTAATACTT GAGGAACATA 3360

TGCCTCAGAT TATAGAAAAA GAAGCTTATT CAATAAATGA TATGTTGTTG ATTCGTTTGT 3420

TTTTTTATCA AATGCTCATT AGAAAAGATC TTGCCAAATT TATAAATCAA ATCGAAAAGC 3480

TAATGCTCTT TCTTTTGGAA CAGAAGAAGG TAACTCAAAT AGAGAATTAC TTTATAATTA 3540

GAGATACTCT TATTTCAGGA ATGTGTTGTC TTGAAAAGGT AGGAGTAACT GATTGTTTTA 3600

ATGATTATCT ATCGTGTTTA CAAGAAATTA TGGATAAAAC TCAAGATTAT CAAAAGAAAC 3660

CTCTTGTATT TATGTTTTTG TGGAAGCAAG CATTAAGAGA AGAAAGAGAT TTTAGTTTAG 3720

CTGAATCATT TTATCAGTCT TCTAAAACAT TTGCGCAGCT AATTGGAGAT GAATTTCTAG 3780

TAAAGAAATT GACAGAGGAA TGGCAAGAGG ATGTCAAAAA ATATTTATAA ACATAGTGAA 3840

TCAGTGACAA AGATGTCCTT GTCCTCGTAT CAAAACAGTT CTAAAGTTCG TCTTTAGGGA 3900

TGTTTTTTTA GATATAAGCT AAAAATGACA CGAAATGGTT AGATTTTAAG GACATTGATG 3960

TCCG 3964

(2) INFORMATION FOR SEQ ID NO: 137:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12666 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

TGAGACCGTT ATTTGTATTA GGGAAATGGG TATCTATTTT TAATGCTGTG GGGATTTTGA 60

TTGTTTCTAT TATTCAAACC AAAAGCTTGT CAGGTATTGG AGCAGGATTG TTTAATCTAT 120

ATAACATTTC ATCTTATATA GGTGATTTAG TTAGTTTCAC TCGATTGATG GCATTAGGAT 180

TATCTGGAGC AAGTATAGCA TCAGCTTTCA ATTTAATTGT TGGTTTGTTT CCGGGAATAT 240

TGGCTAAACT GACAATTGGA TTAGTATTAT TCATTCTTTT ACATGCGATC AATATTTTTC 300

TATCGTTACT ATCAGGATAT GTTCATGGAG CACGTCTGAT ATTTGTTGAA TTTTTTGGTA 360

AGTTTTATGA GGGTGGAGGA AAACCATTTC AACCTTTGAA GGCTTCTGAG AAATATATTA 420

AGGTTATTAC AAAGAATTAA TGGAGGATAT ATATAATGGA ACATTTAGCA ACTTATTTTT 480

CAACCTATGG AGGAGCTTTC TTCGCTGCAT TGGGAATTGT ATTGGCGGTT GGATTAAGCG 540

GTATGGGGTC TGCTTATGGA GTTGGTAAGG CTGGGCAATC TGCCGCAGCT TTACTGAAAG 600

ATTGAAGAGA GGAGGGGAAA CCGAAAAATT AGGTGCCCCT CCTCTTTTTT GGTATAATAG 1500

AAGATAGAAA ACGAGGTTAG AAGAGATGAT TTTTGATACA CATACACACT TGAATGTAGA 1560

AGAATTTGCA GGTCGTGAGG CAGAAGAAAT TGCCTTGGCT GCTGAGATGG GTGTGACACA 1620

GATGAATATT GTTGGTTTTG ATAAACCGAC GATTGAGCAT GCCTTGGAGT TGGTAGATGA 1680

GTATGAGCAG CTCTATGCGA CTATTGGTTG GCATCCTACA GAAGCTGGTA CTTATACAGA 1740

GGAAGTTGAG GCTTACTTGT TGGATAAGTT AAAACATTCC AAGGTTGTGG CTTTAGGTGA 1800

AATTGGCTTA GATTACCATT GGATGACAGC GCCCAAAGAG GTGCAGGAGC AGGTTTTTCG 1860

CCGTCAGATT CAGCTATCTA AGGACTTGGA TTTGCCTTTT GTTGTCCATA CCCGTGATGC 1920

GCTGGAAGAT ACCTATGAGA TTATCAAGAG TGAGGGCGTT GGTCCTCGTG GTGGTATCAT 1980

GCATTCATTT TCAGGGACGC TTGAGTGGGC AGAGAAGTTT GTGGATCTTG GTATGACCAT 2040

TTCCTTCTCA GGAGTGGTGA CTTTTAAGAA GGCAACTGAC CTCCAAGAAG CAGCTAAAGA 2100

GTTACCTTTG GACAAGATGT TGGTGGAAAC AGATGCGCCT TACTTAGCAC CTGTACCCAA 2160

GCGTGGTCGT GAAAATAAAA CAGCCTATAC TCGCTATGTG GTCGACTTTA TCGCTGACTT 2220

GCGTGGTATG ACGACAGAAG AGCTGGCGGT AGCAACGACT GCAAATGCAG AACGAATTTT 2280

TGGACTGGAC AGCAAGTAAT GAAAGAGAAA ATTTCTCAAG TTATCGTGGT TGAAGGGCGT 2340

GATGATACGG TCAATCTCAA ACGTTATTTC GATGTGGAGA CCTATGAGAC TCGAGGTTCT 2400

GCCATCAATG CTCAGGATAT AGAGCGGATT CAGCGCCTGC ACCAACGTCA TGGAGTCATT 2460

GTCTTTACAG ACCCAGATTT TAATGGGGAA CGGATTCGGC GCATGATCAT GATGGTCATT 2520

CCAACAGTTC AGCATGCCTT TCTCAAGCGA GATGAAGCTG TTCCCAAGTC CAAGACCAAG 2580

GGGCGTTCTC TGGGAATTGA GCATGCCAGC TATGAAGACC TGAAAACGGC TCTAGCTCAA 2640

GTGACAGAAC AATTTGAACA TGAGAGTCAG TTTGACATTA GTCGTAGCGA TTTGATTCGC 2700

CTTGGTTTTC TAGCAGGGGC AGACAGCCGT AAGCGTAGAG AATATCTCGG AGAGACTCTC 2760

CGAATCGGCT ATTCCAACGG CAAGCAACTC CTCAAACGCC TAGAGTTGTT TGGGGTTACT 2820

TTGGCAGAAG TGGAAGAAGC TATGAAATCT TATGAGTAGG AAAGATGTAG CCGTTACAAT 2880

TTTTTAAGTT TCACAGTATT TTTCGAAGCA GGTAGAAGAG GAGGCGTCTG ATGTTAATTG 2940

GTCAAAAAAT TAAAGAGATT CGGATAGAAA AAGGAATTAG TCGTCCAGAT TTTTGTGGAG 3000

ATGAGCAAGA ACTGACAGTT CGTCAACTGT CGCGAATTGA AAGTGGAGCT TCGCAACCGA 3060

GTTTGCCCAA GTTAGACTAT ATTGCTCGCC GGCTAGGAGT TCCAGTTTAT AGCCTTATGC 3120

CGGATTTTTC AGCTCTTCCT TCTGCTTATT TAGAATTGAA ATACCAGATT TTACGTGAAC 3180

CAATCTATGG TAAAGAAGAG GAGTACGATA AGAAGGAAGC GTGTTTGGAA GAGATTTATA 3240

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3964 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

TGGCAGCTCG TCGTCGTAAA GGACGCAAAG TTTTGGCTGC ATAATCCAAA CGAATTCTAT 60

CAAAAATCAG TAGGAACTCG AGTCTACTGA TTTTTATTTT TGTAAAAAAG TTCAGTAGAT 120

GCAAATGGAT TCGGAAGCGA TGTTACAGTA GATTGAAACT AGAATAGTAC ACCTCTGTTT 180

CTAAAACATT GTTAGAAATC GATTTGACTG TCCTGATCGA TTTGTCCTGT TATTATTTTA 240

TTTTACTATA AAGTTGAAGT AGGTGGAGAT GGTACAGCAA CAATCGTCTT TAAAGATGGT 300

TCAGCTATTA CAATTCCAGG AAATCAATTG GTAGCACAAG ATCCAAAAGC ACAAGATAGC 360

ACTAAACTGA CTGCTGAAAA ATCAACTGTT AAAGCACCTG CTCAAAGAGT AGATGTAAAA 420

GATATAACTC ATTTAACAGA TGAAGAAAAA GTTAAGGTTG CTATTTTACA AGCAAATGGT 480

TCAGCATTAG ACGGAGCGAC AATCAATGTA GCTGGAGATG GTACAGCAAC AATCACATTC 540

CCAGATGGTT CAGTAGTGAC GATTCTAGGA AAAGATACAG TTCAACAATC TGCGAAAGGT 600

GAATCTGTAA CTCAAGAAGC TACACCAGAG TATAAGCTAG AAAATACACC AGGTGGAGAT 660

AAGGGAGGCA ATACTGGAAG CTCAGATGCT AATGCGAATG AAGGCGGTGG TAGCCAGGCG 720

GGTGGATCAG CTCACACAGG TTCACAAAAC TCAGCTCAAT CACAAGCTTC TAAGCAATTA 780

GCTACTGAAA AAGAATCAGC TAAAAATGCC ATTGAAAAAG CAGCCAAGGA CAAGCAGGAT 840

GAAATCAAAG GCGCACCGCT TTCTGATAAA GAAAAAGCAG AACTTTTAGC AAGAGTGGAA 900

GCAGAAAAAC AAGCAGCTCT CAAAGAGATT GAAAATGCGA AAACTATGGA AGATGTGAAG 960

GAAGCAGAAA CGATTGGAGT GCAAGCCATT GCCATGGTTA CAGTTCCTAA GAGACCAGTG 1020

GCTCCTAATG CTGCTCCTAA GACAACAAGT GCACCGCAAG CAACTGCAGG AACAATGCAA 1080

GATGTTACCT ACCAGTCACC TGCTGGCAAA CAATTACCTA ACACAGGTTC AGCATCAAGT 1140

GCAGCACTTG CTAGTCTTGG TCTAGTGGTG GCAACAAGTG GTTTTGCTTT GCTAGGAAGA 1200

AAGACTAGAC GTAGAAAATA GAACAGCTAG AAAATTCTAT TCTCTACTTA AAGTTAGATT 1260

ATAAGGGGGA TTTTGAGAAG TCATCAATCC TAGTGATGGG TGAGAAAAGT GAGAACCCAA 1320

GATAATCACA TACTTTAGCT GAATAGGAAT ATTCTATCAA TGTAGCCAAT CTCTTCTGTC 1380

TCTAACTGTG GAATAGGAGA TGGGCAATAT CGGATAGAAA AGATAGCAGA ATAGCTCTCT 1440

GGAGGTTTTC AGCCTCTCAT CTTGAAATAA GAAAGTGAGA GAAGGTCTGG GGGATCTTGA 3660

ACCCCGAGTT TAGAAATAAG AAAATGAGGC AGATTCAGTA ACTCGAAGAG TTCGATTTCA 3720

TCGTCTTACC CCTGCAACGA TGACTAGGTT TGAAAAAGCT TGCTAGAGCG CATTTCAAAC 3780

CAGGCAGCAA CTGCGTCAAG AAATTAGAAG ACAAACTCGT TTTCTAGCTG TTACTGAGTT 3840

GAGCCTTTTT ACTACGAGTA TAGAAATAAG GAAGTGAGGT AGCATCATGA AATCTATCGG 3900

TACGCAAATA TTACAGACAG AACGTTTGAT TTTAAGAAGA TTTGTGGAGA GTGATGCAGA 3960

AGCCATGTTT CAAAATTGGG CTTCATCCGC TGAGAATCTG ACCTATGTTA CCTGGGATCC 4020

CCATCCTGAT GTCGAAATCA CTCGAAACTC GATTTGCAAT TGGGTTGCTT CCTATACTAA 4080

TCTCAACTAT TATAAATGGG CCATTTGTCT AAAAGAAAAC CCAGAGCAAG TAATAGGAGA 4140

TATCAGCATT GTTAAGATAG ACGAGGCTGA TTTAAGCTGT GAAATTGGCT ATGTGTTAGG 4200

CAAGGCTTAC TGGGGAAATG GTATGATGAC AGAGACTTTG AAAGCTATCT TGGACTTTTG 4260

TTTTACTCAA GCAGGTTTTC AAAAGGTCAG AGCACGTTAT GCCAGTCTCA ACCCAGCTTC 4320

AGGTCGTGTC ATGGAAAAGG CTGGAATGTC CTATCTACAA ACCATTGTTA ATGGTGTAGA 4380

GAGAAAAGGC TATCTTGCGG ATCTTATTTA TTATGGTATA AGTAGGGAAG AATGTTGAAT 4440

TCTATTTTCT GTTTCTATCG AAGTCAACTA TTTATTGTAA ATATAATAAT TAGCATTCCA 4500

AGTTTATTTG AAACTTTAAA ATAGCATATT GATTAGTACA AGACAGATGT TCTAGTTCCT 4560

TCTTTAATCT GGTTTAGTGT TAGTTAAAAA ATCGCTTTAA GCTTGTAACT AAGAGGGAGC 4620

TAATCGACTA GATTCTCCAG CCGAACAGGT GGTAATGTAC TTTTTATAGT GTAATCCTAG 4680

CTGTTGTTAA ATTTAAAATA GAATCCTCTA TCGAGTTAGG GAATTAAATT CAACCAATTT 4740

TATTCATGTT TTTTCTATCA AATTATCTAA TATTAAAATA GTCTCATTCT GATGAGAAAA 4800

CTATTCCCAA ATCATTCATA CCTCTCTCAA CTAGATGTAA CTTACAAAAC CCCTGACCTC 4860

ATGAGCCACT TTCTTCCTCC TCATGAGGTC AGTTTTACTT TCTGCTGTTC CAGTATCGTT 4920

TTTCCTCGCT AGATTTCCTC AAAAGGGCAG ACTCCTCCCT TGGTGCGTCA CACGATTTTT 4980

TCATCTCGAC TGTTCTTTAA TGCATCATTA ACGACGCTTT TCTTCTAGGT GGTTCATAAG 5040

GAACAGGAAG ATTCAGGTTG ACTTTTCTAA TCCTAGAATA AAGTGCTGAA AACAATTCGG 5100

AATAGGCATA GAGACTAGAC AATTTGAGGA GCTGCTTGCG TCCTGTTCGA ACACATTTTC 5160

CCACCACGTG AAGAAAAAGA TGGCGGAAGC GTTTGATTGT TAAAGTTTGG AAGTCACCTC 5220

CAGCTAGATG TTTGAGAAAA AGATAGAGAT TGTAGGCGAT ACAGCTCATC ATCATACGAA 5280

CTTCGTTTTT GATTAAGGTT GAACT 5305

(2) INFORMATION FOR SEQ ID NO: 136:

CTTGGTACAG GTTTATGGCA ATGACTATGA AGCTTTCAAG GCTTACGCTG CGACCCACAA 1920

AAATTGTGTC TTTCTTGTGG ATACCTATGA CACCCTTCGC ATCGGTGTAC CAGCTGCCAT 1980

TCAGGTGGCG CGTGAGCTGG GTGATCAGAT TAACTTTATG GGTGTGCGGA TTGACTCTGG 2040

GGATATTGCC TACATTTCTA AGAAAGTCCG TCAGCAACTG GATGAGGCTG GATTTACAGA 2100

GGCTAAGATT TATGCTTCTA ATGATCTAGA TGAAAATACC ATCCTTAACC TCAAGATGCA 2160

AAAGGCCAAG ATTGATGTCT GGGGTGTGGG TACCAAGCTG ATTACAGCCT ATGACCAGCC 2220

GGCTCTTGGG GCGGTTTACA AGATTGTTGC AATCGAAGAT GAAACTGGTC AGATGCGCAA 2280

TACGATTAAG CTGTCTAATA ATGCTGAAAA AGTTTCTACG CCAGGTAAGA AGCAGGTGTG 2340

GCGCATTACC AGTCGTGAAA AAGGCAAGTC AGAAGGCGAC TATATCACTT ATGATGGTGT 2400

GGATATTAGC GACATGACAG AAATCAAGAT GTTCCATCCG ACCTATACAT ACATCAAGAA 2460

GACGGTTCGT AATTTTGATG CCGTTCCTCT CTTGGTGGAT ATCTTCAAAG AAGGAATATT 2520

AGTTTACAAC TTGCCTAGTT TGACTGACAT TCAGGATTAT GCCCGTAAGG AATTTGACAA 2580

GTTGTGGGAT GAGTATAAGC GTGTGCTCAA TCCGCAGCAC TATCCAGTGG ATTTGGCGCG 2640

TGATGTATGG CAAGATAAGA TGGACTTGAT TGATAAGATG CGCAAGGAAG CCCTTGGTGA 2700

AGGAGAAGAA GAATGAGTTT GCAAGAAACG ATTATCCAAG AGCTGGGTGT CAAACCAGTG 2760

ATTGATGCCC AGGAAGAAAT CCGTCGTTCT ATTGATTTCT TAAAAAGATA TCTGAAAAAA 2820

CATCCCTTCC TAAAAACCTT TGTACTAGGG ATTTCTGGGG GACAAGACTC AACCTTGGCA 2880

GGACGTTTGG CGCAATTAGC TATGGAAGAA CTGCGAGCTG AAACGGGAGA CGATAGCTAC 2940

AAATTTATCG CTGTCCGCCT GCCATACGGA GTGCAAGCTG ATGAAGCAGA TGCTCAAAAA 3000

GCCCTAGCCT TCATCCAGCC AGATGTCAGC TTGGTTGTGA ATATCAAGGA ATCAGCTGAT 3060

GCCATGACAG CTGCAGTTGA AGCGACAGGT AGTCCTGTTT CAGACTTCAA CAAGGGGAAT 3120

ATCAAGGCAC GTTGCCGTAT GATTGCTCAG TATGCCCTTG CTGGTTCCCA TAGCGGAGCG 3180

GTCATTGGAA CAGACCACGC CGCGGAAAAT ATCACAGGTT TCTTTACCAA GTTTGGTGAC 3240

GGCGGTGCGG ATATTCTCCC TCTTTACCGC CTCAATAAAC GCCAAGGAAA ACAGCTCTTG 3300

CAGAAACTTG GCGCAGAGCC AGCCCTTTAT GAAAAAATCC CAACGGCAGA CCTAGAAGAA 3360

GATAAACCAG GCCTAGCTGA CGAAGTCGCA CTTGGAGTCA CCTACGCAGA GATTGACGAC 3420

TACCTAGAAG GCAAAACAAT CAGCCCAGAA GCTCAAGCGA CCATTGAAAA CTGGTGGCAC 3480

AAAGGCCAAC ACAAACGCCA CTTACCCATC ACCGTATTTG ATGACTTTTG GGAGTAAAAA 3540

GGTCCGGGGG ACCTTTTTAG CTTCTTGCCC TGAAATTAAA AAGCAAGAAA AACCTCCACT 3600

CATTTTAGTT TCATTTAGTA TTATTTTGCA TACCTAAAAT ACAGTAAAAA ATCAGTCATC 120

TTGGTATGCT CCTGCTTTCA CTATTCAACA CGTTTTTGAC TTATACTAGG CTCATTTCCA 180

AAAGCATTAT ATAATAGTGA TATGAAACCA ACTAAACTAA ACAAGAAATA TAAGCAATAA 240

AAATTCGTTT AAAAGATCTT ACTAAAGCTA ATACTAAATA AAAATAAAAG AGTAAACTAG 300

GAAGTTTATT TCAAACAACC TAAAATACTG ATTTTCGGCT GAAGATAATA CTGGAGTGCA 360

AATTAATGGG GTTATAATAA ATAGCTGATA GCTTGTGTTG GTTTTGGATT TTTTAAGAGT 420

AGATGAGTAT TAAAACTATA AGGAGGACGA AGGTGGCTAA AAATTTAAAA TTAAAATTAG 480

CTCGGGTAGA GCGTGATTTA ACACAAGGTC AACTGGCAGA GGCTGTCGGG GTGACACGCC 540

AGACTATTGG TTTAATAGAG GCGGGAAAAT ACAATCCCAG TCTCTCGCTC TGCCAGTCTA 600

TTTGCAGATG TTTAGGGAAA ACCCTAGACC AACTATTTTG GGAGGAAGAA GATGAAAAAT 660

AGATTTTATT ATTCTCAATT ACTAGACGAA AGAGAAGAAC AACTGTTCAA TAAAGCGGGC 720

TCTGAAAGTT TCTATATCTG CATTGCTTTG TCGCTCCTAT CTTATATCAT TTCAGTATTA 780

GCACCAAGCC TTTTTAATTC TAATATGCTG CTAATCGTTA TCATCATAGG GACATTTTAC 840

TTTTTCAATC GTGCCCGTTA TCTGGGAGTG ACCTACTATG GTCGTTTTCA TTTTACGATT 900

TTGGGTTGTT TTTTCCTAAC CTTGGCTATT ACGGCTCTTT TGATGTTGCA GAATTATCAA 960

TTCAACATAG AAATTTATCA GCACAATCCT TTGAATTTTA AATACCTGTC TGCTTGGGTC 1020

ATTACTTATA TCATTTACCT TCCGTGGATC TTTATTGGCA ATCTTGGTCT TAAGAGCTAT 1080

GGCGAATGGG CTCAGAAAAA ATTTGAACAA GATATGGATG AATTGGAGAG TGGAGAATAG 1140

CTTGTTACTC TTTTCTCAAT CCAGCTAAAA TGTGATATAA TAGTACTAAT TTATTGGAAT 1200

ACATGAAAGT TCTTGAAAAT TTTCATGGGT TTCTAGCTAA GGAAGTAGGA AAAGTATGTA 1260

TCCAGATGAT AGTTTGACAT TGCACACGGA CTTGTACCAG ATCAACATGA TGCAGGTTTA 1320

CTTTGACCAA GGGATTCACA ATAAGAAGGC GGTCTTTGAG GTGTATTTCC GCCAACAGCC 1380

TTTTAAGAAC GGCTATGCGG TTTTTGCAGG TTTAGAAAGA ATTGTGAACT ATCTTGAAGA 1440

CTTGCGTTTT TCAGATAGTG ATATAGCCTA TTTGGAGTCG CTTGGTTATC ATGGGGCGTT 1500

CTTGGATTAC CTTCGCAATT TCAAGTTGGA GTTGACCGTT CGTTCTGCCC AAGAAGGGGA 1560

TTTGGTTTTT GCTAATGAAC CGATTGTGCA GGTGGAAGGA CCTCTAGCCC AATGTCAGTT 1620

GGTCGAAACG GCTCTTTTGA ACATCGTCAA CTACCAGACT TTGGTGGCGA CGAAGGCAGC 1680

TCGTATTCGT TCGGTTATCG AAGATGAACC CTTGATGGAG TTTGGGACAC GTCGGGCTCA 1740

AGAAATGGAT GCGGCCATCT GGGGAACACG CGCAGCTGTG ATTGGTGGCG CCAATGGAAC 1800

CAGCAACGTG CGTGCGGGTA AGCTCTTTGA CATTCCTGTT TTGGGAACCC ATGCCCATGC 1860

TGCACCCCAT GGACCCCCCA ATAAAGCACC TATCCTACCA ATCATATAAC TGATTCCAGC 11460

ACCAGTCATG AAGTTAGCGA ATGTGTTAGC TTGTTTATTC CCATGTATTG TGTTGACGTA 11520

ATTCCAAACA TTAGGATCGT ATGATCTAAA AGATATATTT AGGTCGATTT CATTCTTTTG 11580

ATAAGCCATA TAAAATGCCC CATTGATATA GACGCCGTCA GCACGTCGTT CAATAGTGTC 11640

TACACTTCCA TCTGGATTGA CAACCTCAAG AACTTCATCG CTTAAAATAT TTACTTGCGT 11700

ATCTCCGAAC CGCACTGATG AGCCATTCTC AAACTGAGCC TCACCAGATA CAACTTTAGA 11760

GTTTGCCGAT AAGCTATCAT CAGCAAAAAC AAACAAGCGA CGGGGAAATG CTAGACATAC 11820

AGAAAACAGA CATAACTAGC AAACACATGC ATTTAAACAT CTTAGACATA ACGGAAACTC 11880

CTTTGTATTT TTGATTTTTT TCAACTTTTA TTATACAATA AAACCAAATA AAAAGAAAGC 11940

GGTAACAATA TGCTTAATGC GAAAATTTTT TATATATTTT TATGTTTGAT CGTTATCGAA 12000

ACTACAGGCT TGTTGTTGTT GAAAAGAGGT CTCGAAATGG GTTATTTAGA CACAGAAGCT 12060

ATTATCCTCG CAGTTTTTTC ATTTGCTTTT TACAACCTAT GTTCATTCGC TTGGGTCTGC 12120

TCTACAATAA AAAACAATAA AAAATAAATA GACGTATTTT CAAAAAAAAC maAATGCATA 12180

TTTATATTAG CAAAACGACG ATTTAAATCG TCGTTTTTTT GTAGTACGAC GGGCATGTCG 12240

TATATCTGAG GTGTAAGTCC TCAGCCTGAC TATCGTGAGG TAGCAGGGAG AGGAAGGGAT 12300

AGCGAAATCG TGGCTCTACG AACAGGAACG TGATAGTAAG GCGTATATAG CGGATAAGGA 12360

GGCTTCAAAC TCTAAAGTCC AAAAAGGTAG TCGTAACCTA TATGTGTAAA TCACGAGAGT 12420

AATTGAATTC GGACTAAGGT TTGTGTGAAA AAGATAAATC TTTCTAGAGT CTAAAGACTC 12480

TGCGTCAGAT TTCCTATTTT CACTGTAACC TTTTAACGTC CTCATATCTT GTATAAACGA 12540

GGAAAGATGT ACGACTTATC CCGTGAGGTT TCATGAGCGT GAAAGCGTAG TAACAACGAA 12600

TCATGAGAAG TCAGCCGAGC CCATAGTAGT GAGGAAACTT CCGTAATGGA AGTGGAGCGA 12660

AGGGG 12665

(2) INFORMATION FOR SEQ ID NO: 135:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5305 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

CGCTAATCAC TACAATCATT TTATTGTACT TTTTCACTCT CAAGAAAAGC AAGAAGTATT 60

TCAGCTTGGC AGATACTTTG TCCCTTCACT ATGCAATGCG CAATGAGTTG ACTCTTAGCC 9660

CAGTTGACTT CCTTCTTCGT CGTACCAATC ACATGCTCTT TATGCGTGAT AGCTTGGATA 9720

GTATCGTTGA GCCAATTTTG GATGAAATGG GACGATTCTA TGACTGGACA GAAGAAGAAA 9780

AAGCAACTTA CCGTGCTGAT GTCGAAGCAG CTCTCGCTAA CAACGATTTA GCAGAATTAA 9840

AAAATTAAGA AAAAATAAAA GAGGTGGAGG GCAGCATTCC TTGTCGCCCG TCCCTTCTTT 9900

TTAATGGAGA CAGAAAGATG ATGAATGAAT TATTTGGAGA ATTTCTAGGG ACTTTAATCC 9960

TGATTCTTCT AGGAAATGGT GTTGTTGCAG GTGTGGTTCT TCCTAAAACC AAGAGCAATA 10020

GCTCAGGTTG GATTGTGATT ACTATGGGTT GGGGGATTGC AGTTGCGGTT GCAGTCTTTG 10080

TATCTGGCAA GCTCAGTCCA GCTTATTTAA ACCCAGCTGT GACCATCGGT GTGGCCTTAA 10140

AAGGTGGTTT GCCTTGGGCT TCCGTTTTGC CTTATATCTT AGCCCAGTTC GCAGGGGCCA 10200

TGCTGGGTCA GATTTTGGTT TGGTTGCAAT TCAAACCTCA CTATGAGGCA GAAGAAAATG 10260

CAGGCAATAT CCTGGCAACC TTCAGTACTG GACCAGCCAT CAAGGATACT GTATCAAACT 10320

TGATTAGCGA AATCCTTGGA ACTTTTGTTT TGGTGTTGAC AATCTTTGCT TTGGGTCTTT 10380

ACGACTTTCA GGCAGGTATC GGAACCTTTG CAGTGGGAAC TTTGATTGTC GGTATCGGTC 10440

TATCACTAGG TGGGACAACA GGTTATGCCT TGAACCCAGC TCGTGACCTT GGACCTCGTA 10500

TCATGCACAG CATCTTGCCA ATTCCAAACA AGGGAGACGG AGACTGGTCT TACGCTTGGA 10560

TTCCTGTTGT AGGCCCTGTT ATCGGAGCAG CCTTGGCAGT GCTTGTATTC TCACTTTTCT 10620

AGTTTATACT CTTCGAAAAT CAAATTCAAA CCACGTCAGC GTCGCCTTAC CGTACTCAAG 10680

TACAGCTTGC GGCTAGCTTC CTAGTTTGCT CTTTGATTTT CATTGAGTAT TAGAAAACAA 10740

TTATGTTGAT AGAGCTTGGG CAAGAGCCCA ATTTCAGCAA AAAATGAAGT AAATCTTCTC 10800

ATAATAAAAC GCATCATATC AAGCACGAAA ATTCCACGAG GTCAACTACA GTCAGAAAGC 10860

TGAACAACAA GCCAAAACGC CCAAAAAAGG CGGCAAAAAG CAAGCACCTG CAAGCAACGT 10920

GCCGAAATGG TCAAATCCTG ATTATGTCAA CGAATTAGAC CCAAAAATCG TTGATATGCT 10980

AGTAGAATTT CACAAGTCAC AAGGCACTTT GGAAACTCCC GAGGCGCAAG CAGAAATCGC 11040

CCAAAAACGT GAAGAAATCG AGCAAAGGAG AGCTGAGCTT GAGGGTAAAA AACAAGAGCT 11100

TTTGAACCGC TTGAACAAAT AGAGTTTCGC AAGTATTATG CTTACAAATT ACTTGAGCAA 11160

TTAACTAAAA TATAAACCCT GCCTTTATAT CTAGGCAGGG TTTATATTTT AGAAATTCAC 11220

GTAGGTTGTT ACGGTTTTTA CATACCCAGT ATAGTTTGAG TTTCTATAGT ATTCAGTGAT 11280

AAACTTCCAT TTTCTTTGAG CAACATGGAT ATAAGTACTT GTTATGTAGT ATGGATATGG 11340

GCTTTGTGAA TCCAAGTAAG ACTGATAAGC TTGTATACCA AAATATGCTC CACCAATTAT 11400

TTGAACGAGA CAGGAGAACT CTTTGAGCCA TCTATGAACG AATCTCGCAA GGAACAACTC 7920

TACAAGGGCT GGAAGAAGGC TGTGAAAGCA ACTCAAGTCT TTGCGGAAGT AGACGACTAA 7980

TACTGGCAGA ATAAAGCGAT TTATTTAGAA AGTGTGTAAA TATGGAATTT TCAAAGAAAA 8040

CACGTGAATT GTCAATTAAA AAAATGCAGG AACGTACCCT GGACCTCTTG ATTATCGGTG 8100

GAGGAATCAC AGGAGCTGGT GTAGCCTTGC AGGCGGCAGC TAGCGGTCTT GAGACTGGTT 8160

TGATTGAAAT GCAAGACTTT GCAGAAGGAA CATCTAGTCG TTCAACAAAA TTGGTTCACG 8220

GAGGACTTCG TTACCTCAAA CAATTTGACG TAGAAGTGGT CTCAGATACG GTTTCTGAAC 8280

GTGCAGTGGT TCAACAAATC GCTCCACACA TTCCAAAATC AGATCCAATG CTCTTACCAG 8340

TTTACGATGA AGATGGAGCA ACCTTTAGCC TCTTCCGTCT TAAAGTAGCC ATGGACTTGT 8400

ACGACCTCTT GGCAGGTGTT AGCAACACAC CAGCTGCGAA CAAGGTTTTG AGCAAGGATC 8460

AAGTCTTGGA ACGCCAGCCA AACTTGAAGA AGGAAGGCTT GGTAGGAGGT GGAGTGTATC 8520

TTGACTTCCG TAACAACGAT GCGCGTCTCG TGATTGAAAA CATCAAACGT GCCAACCAAG 8580

ACGGTGCCCT CATTGCCAAC CACGTGAAGG CAGAAGGCTT CCTCTTTGAC GAAAGTGGCA 8640

AGATTACAGG TGTTGTAGCT CGTGATCTCT TGACAGACCA AGTGTTTGAA ATCAAGGCCC 8700

GTCTGGTTAT TAATACAACA GGTCCTTGGA GTGATAAAGT ACGTAATTTG TCTAATAAGG 8760

GAACGCAATT CTCACAAATG CGCCCAACTA AGGGAGTTCA CTTGGTAGTA GATTCAAGCA 8820

AAATCAAGGT TTCACAGCCA GTTTACTTCG ACACAGGTTT GGGTGACGGT CGTATGGTCT 8880

TTGTTCTCCC ACGTGAAAAC AAGACTTACT TTGGTACAAC TGATACAGAC TACACAGGTG 8940

ATTTGGAGCA TCCAAAAGTA ACTCAAGAAG ATGTAGATTA TCTACTTGGC ATTGTCAACA 9000

ACCGCTTCCC AGAATCCAAC ATCACCATTG ATGATATCGA AAGCAGCTGG GCAGGTCTTC 9060

GTCCATTGAT TGCAGGGAAC AGTGCCTCTG ACTATAATGG TGGAAATAAC GGTACCATCA 9120

GTGATGAAAG CTTTGACAAC TTGATTGCGA CTGTTGAATC TTATCTCTCC AAAGAAAAAA 9180

CACGTGAAGA TGTTGAGTCT GCTGTCAGCA AGCTTGAAAG TAGCACATCT GAGAAACATT 9240

TGGATCCATC TGCAGTTTCT CGTGGGTCTA GCTTGGACCG TGATGACAAT GGTCTCTTGA 9300

CTCTTGCTGG TGGTAAAATC ACAGACTACC GTAAGATGGC TGAAGGAGCT ATGGAGCGCG 9360

TGGTTGACAT CCTCAAAGCA GAATTTGACC GTAGCTTTAA ATTGATCAAT TCTAAAACTT 9420

ACCCTGTTTC AGGTGGAGAA TTGAACCCAG CAAATGTGGA TTCAGAAATC GAAGCCTTTG 9480

CGCAACTTGG AGTATCACGT GGTTTGGATA GCAAGGAAGC TCACTATCTG GCAAATCTTT 9540

ACGGTTCAAA TGCACCGAAA GTCTTTGCAC TTGCTCACAG CTTGGAACAA GCGCCAGGAC 9600

TGGCCAGCAT ATGCGGATTG GTCTGGATTT GACATCTGGT TTTCTTGTCT TTTCAAGGAT 6120

GGCAGCCATT TTGAAACGGT ATTTGGAATA CAATCGTTTT ATTACCATTG AAGCTTATGA 6180

CCCTAGTCGG CATTATGATT TGCTGGTTAC CAATAACCCG ATTCATAAGA AGGAACAGAC 6240

ACCAGTCTAT TATTTAAAAA ATGACTTGGA TATGGAGGAT TTGGTAGCGA TTCGCCAGTT 6300

ATTATTCACT TAAAAGGCTT GGTTAATCCA GGTCTTTTTT GTGAAATTCA CACAATCTCC 6360

TCACATTTTT TTAAAAATTA AAAAAAGTTG ATAAACAAGA AAGCGCTTTA TTTTGTATAC 6420

TAGTAAGTGT AAAGAGGAAA CACCTCAAGA TCTTTATCAG GAGGACAGTA CATGTCACAA 6480

GAAAAATACA TCATGGCCAT TGACCAGGGA ACTACAAGTT CTCGTGCCAT CATTTTCAAC 6540

AAAAAAGGGG AAAAGGTTAG CTCGAGTCAA AAAGAGTTTA CCCAGATTTT CCCTCAGGCA 6600

GGTTGGGTTG AGCACAATGC CAATGAAATT TGGAACTCTG TTCAGTCAGT TATTGCGGGT 6660

GCTTTCATCG AAAGTGGTGT CAAGCCAAAT CAAATCGAGG CAATCGGGAT TACCAACCAA 6720

CGTGAAACAA CGGTTGTCTG GGATAAGAAA ACAGGACTTC CTATCTACAA TGCTATCGTT 6780

TGGCAGTCAC GCCAGACAGC ACCTTTGGCT GAGCAACTAA AAAGCCAAGG TTATGTGGAA 6840

AAATTCCATG AAAAGACTGG TTTGATTATT GATGCTTACT TCTCTGCTAC CAAGGTTCGT 6900

TGGATTTTGG ATCATGTAGA AGGTGCTCAA GAGCGAGCAG AAAAAGGGGA ATTGCTCTTT 6960

GGTACTATCG ATACTTGGTT GGTTTGGAAA TTGACTGACG GTGCGGCTCA CGTGACTGAC 7020

TACTCAAATG CAGCTCGTAC CATGCTTTAT AACATTAAAG AACTCAAATG GGATGATGAG 7080

ATTTTGGAAA TCCTTAACAT TCCGAAGgCT ATACTTCCAG AAGTTCGTTC TAACTCCGAA 7140

ATCTACGGCA AGACAGCTCC ATTCCATTTC TACGGTGGAG AGGTGCCAAT CTCAGGTATG 7200

GCTGGGGACC AACAAGCAGC CCTCTTTGGA CAGTTGGCTT TTGAGCCAGG TATGGTTAAG 7260

AATACTTATG GAACAGGCTC TTTCATCATC ATGAATACTG GGGAAGAGAT GCAGTTGTCT 7320

GAAAACAACC TCTTGACAAC CATTGGTTAC GGAATCAACG GTAAGGTTTA TTATGCCTTG 7380

GAAGGTTCTA TCTTCATCGC AGGAAGTGCT ATTCAGTGGC TTCGTGACGG TCTTCGCATG 7440

GTTGAAAATT CACCAGAATC TGAAAAATAC GCTCGTGATT CTCACAACAA CGATGAAGTT 7500

TATGTCGTTC CAGCCTTTAC AGGTCTAGGC GCTCCATACT GGAACCAAAA TGCTCGTGGT 7560

TCCGTCTTTG GTTTGACTCG TGGAACAAGC AAAGAAGACT TTATCAAGGC GACTTTGCAA 7620

TCTATTGCTT ATCAAGTGCG TGATATCATC GACACCATGC AAGTGGATAC TCAGACCGCC 7680

ATTCAAGTAC TGAAGGTGGA TGGTGGTGCA GCCATGAACA ACTTCCTCAT GCAGTTCCAG 7740

GCGGATATTT TAGGCATTGA CATTGCACGT GCTAAAAACC TGGAAACAAC AGCTCTAGGA 7800

GCGGCCTTCC TAGCAGGTTT GTCAGTAGGG TACTGGAAAG ACTTGGACGA GTTGAAACTC 7860

ATGAGGGGAG TTATAGAGTT GTAAGGATTT CCAGTACCGT AGTCTGTGAT AACGAGGAAT 4380

TGACCATTTC CAACAAAAGG TCCGACTAGG ACTTCACCAG TCGCAGTCTT TTTGATGTCA 4440

ACACCAGGAT TTTGAACATA GCCTTTGACG TTCCCCTTGG TATCAGCGAC GGTGATAATA 4500

GCACCTAGAG AGCTAGATCC CAACACCTTA ACTGTAAGTT TGGTATTTCC TTTTTCATTG 4560

GCTGCGAGAA TCTGGCTAGC GATAAGAGTT CGACCAAGCG CTACAGTTGA GCTAGCTTGG 4620

GTTTGATGTT TTTCTTGAGC AGTGCGGACG GTTTCAGTGC TATCAAGGAC AAAAGCACGA 4680

AAGGCTCCGC TTTCTGATAT AGTTTTAATA ATTTTATCCA TAGCTACTAT TTTAGCATAA 4740

AAATGCCCAA AGGGGGAGCC GTGTGTTTAC TGATTTTCAG GATAATGGAC CAGGAAATCA 4800

GCATGAAAAT AAAAAGAGAA ACAGATTATT TTAGCATTTG TCAGATTTAT GCTATGCTTA 4860

AGGTAGAAAA TGAAAGGGAT AACAAATGTA TTTAGGAGAT TTGATGGAGA AAGCCGAGTG 4920

TGGTCAATTT TCAATACTTT CCTTTCTATT ACAAGAGTCT CAGACGACCG TCAAGGCTGT 4980

AATGGAAGAA ACAGGATTTT CAAAAGCAAC CCTAACCAAA TATGTCACCC TGCTCAATGA 5040

CAAGGCTTTG GATAGTGGCT TAGAGCTGGC TATTCACTCA GAAGATGAAA ATCTGCGTCT 5100

GTCTATCGGT GCAGCTACCA AGGGGAGAGA TATTCGGAGC TTGTTTTTGG AGAGTGCTGT 5160

TAAATACCAG ATTTTGGTTT ATCTTCTCTA CCACCAACAG TTTTTAGCCC ATCAGCTGGC 5220

TCAAGAATTG GTGATTAGCG AGGCTACGCT TGGTCGTCAC TTGGCTGGTT TAAATCAGAT 5280

TTTGTCAGAA TTTGATTTAT CCATCCAAAA TGGCCGTTGG CGAGGTCCAG AGCATCAGAT 5340

TCACTATTTC TATTTCTGTC TTTTCCGAAA GGTCTGGTCG AGTCAGGAAT GGGAAGGTCA 5400

CATGCAGAAA CCAGAGAGAA AACAGGAGAT TGCCAATTTA GAGGAAATCT GCGGTGCAAG 5460

TTTGTCTGCG GGGCAGAAAT TGGACTTGGT TCTCTGGGCT CACATCAGTC AACAACGTCT 5520

TCGGGTCAAT GCTTGTCAGT TTCAAGTCAT AGAAGAGAAA ATGCGAGGGT ATTTTGACAA 5580

TATCTTTTAT CTTCGTTTGC TGAGAAAGGT TCCGTCCTTT TTTGCTGGGC AACATATTCC 5640

ACTAGGAGTT GAGGATGGTG AGATGATGAT ATTCTTCTCT TTTCTCCTAT CTCATCGCAT 5700

TCTTCCTCTT CATACTATGG AGTATATTCT TGGTTTTGGA GGGCAGTTGG CAGATTTACT 5760

GACGCAATTG ATTCAAGAAA TGAAGAAGGA GGAACTATTG GGGGATTATA CAGAGGACCA 5820

TGTCACCTAT GAACTCAGTC AGCTTTGTGC TCAAGTCTAT CTCTATAAGG GCTATATTTT 5880

ACAGGATCGC TACAAGTACC AGTTAGAGAA TCGTCATCCA TATTTACTGA TGGAACATGA 5940

TTTTAAAGAG ACAGCAGAGG AGATTTTTCA TGCTCTACCT GCTTTTCAAC AGGGGACAGA 6000

TTTAGATAAG AAGATTCTCT GGGAATGGCT CCAGTTAATC GAATATATGG CTGAAAACGG 6060

GGTGCTATGA AAGCAAGCCA ATGGTTCAAA GTATCAGATA AATGGTACTA TGTCAATGGC 2580

TCAGGTGCCC TTGCAGTCAA CACAACTGTA GATGGCTATG GAGTCAATGC CAATGGTGAA 2640

TGGGTAAACT AAACCTAATA TAACTAGTTA ATACTGACTT CCTGTAAGAA CTCTTTAAAG 2700

TATTCCCTAC AAATACCATA TCCTTTCAGT AGATAATATA CCCTTGTAGG AAGTTTAGAT 2760

TAAAAAATAA CTCTGTAATC TCTAGCCGGA TTTATAGCGC TAGAGACTAC GGAGTTTTTT 2820

TGATGAGGAA AGAATGGCGG CATTCAAGAG GCTCTTTAAG AGAGTTACGG GTTTTAAACT 2880

ATTAAGCCTT CTCCAATTGC AAGAGGGTTT CAATCTCTGC CAGGGTGCTG GCTTGCGAAA 2940

TGGCTCCACG GAGTTTGGCA GCGCCAGATG TTCCACGGAG ATAGTGAGGA GCGAGACCGC 3000

GGAATTCACG AACTGCGACG TTTTCTCCTT TGAGGTTAAT CAATCGTTTC AAGTGTTCGT 3060

AGGCGATCTT CATCTTGTCT TCAAAGGTCA AATCAGGTAG GATTTCTCCT GTTTCAAAGT 3120

AATGGTTGAT TTGGTTGAAG AGGTAAGGAT TTCCCATGGC AGCTCGGCCA ATCATGACTG 3180

CGTCAGCACC AACTTCTTCG ATGCGTTGCT TGGCTTCTTG GACAGTACGG ATATCACCGT 3240

TGGCGATGAA TGGAATCTTG GTTAGAGCTT GGGCAACCTT GTAAAGGGTC TCAAGGTCTG 3300

CGTGGCCAGT ATACATTTGT TCACGGGTAC GGCCATGCAT GGCGAGGGCA GAAACACCTG 3360

CAGCTTCAGC AGCGAGAGCA TTTTCTACTG CAAGAGATGG GTCCGCCCAG CCGGTACGCA 3420

TTTTGACAGT AAGTGGGATA TCAAGGACAG ACTGGACCTT GTTGATGATG GAGTAAATCT 3480

TGTCTGGATC CTTGAGCCAC ATAGCACCAG CTTCGTTCTT CACGATTTTG TTGACAGGGC 3540

AGCCCATGTT GATATCGACG ATATCGGTCT TGGTGTTTTC TTGGATGAAT TCTGCTGCGC 3600

GTGCTAGGCT GTCTTCATCG CTACCAAAAA GTTGGATAGA GACAGGGTTT TCGCCCTCAT 3660

CGATATGAAG CATGTGCAGG GTTTTTTCGT TGTTGTATTG GATTCCCTTG TCAGAGACCA 3720

TTTCCATTAC AACGAGTCCA GCTCCGAGCT CCTTTGCGAT AGTACGAAAG GCTGAGTTGG 3780

TCACGCCAGC CATAGGCGCT AAAACGGTAC GATTGGGAAT CTCAATATTG CCAATCATAA 3840

AAGGTGTATT AAGATTTGTC ACGAATGAGT TCCTCCAGGT CCTTTTCATC AAAGTTGTAA 3900

GTAGTTTGGC AGAATTGACA AGTGATTTCT GCCCCGTGGT CTTCCTCTTT CATTTCCTGT 3960

AAGTCTGAGC TTGGAAGGCT GGCAAGAGCG TTCATAAAGC GTTCATGGCT ACAGTCACAT 4020

TGGAAACGGA TTTCTTCTTC AGAAAGACGC TTGTAGGCTT CGTCCCCGTA GATAGCCTTG 4080

AGGAGGGCTT CGATATGGTC GTCGCTTTCG AGAAGAGTAG AGATAGCTGG CATTTCTTGG 4140

ATGCGTTTTT CAAAGCGAGC AATCTCTTCT TTCTTGGCTC CTGGCAAGAC TTGAACTAGG 4200

AAACCACCTG CAACCTTGAC CTTGTCTTCC TCGTCCAAAA GGACATTGAG GCCGACCGCT 4260

GAAGGCGTTT GTTGGCTTTC AGTAAGGTAA AAGGCAAGGT CTTCACCGAT TTCTCCAGAG 4320

CGAGATAAGG CAAGGAAAGA GGTCGAGGAA TATGTAAAAA AAATAGTGGG TGAGAGCTAT 840

GCAAAATCAA CTAAAAAGCG ACATACAATT ACTGTAGCTC TAGTTAACGA GTTGAACAAC 900

ATTAAGAACG AGTATTTGAA TAAAATAGTT GAATCAACCT CAGAAAGCCA ACTACAGATA 960

CTGATGATGG AGAGTCGATC AAAAGTAGAT GAAGCTGTGT CTAAGTTTGA AAAGGACTCA 1020

TCTTCTTCGT CAAGTTCAGA CTCTTCCACT AAACCGGAAG CTTCAGATAC AGCGAAGCCA 1080

AACAAGCCGA CAGAACCAGG AGAAAAGGTA GCAGAAGCTA AGAAGAAGGT TGAAGAAGCT 1140

GAGAAAAAAG CCAAGGATCA AAAAGAAGAA GATCGTCGTA ACTACCCAAC CATTACTTAC 1200

AAAACGCTTG AACTTGAAAT TGCTGAGTCC GATGTGGAAG TTAAAAAAGC GGAGCTTGAA 1260

CTAGTAAAAG TGAAAGCTAA CGAACCTCGA GACGAGCAAA AAATTAAGCA AGCAGAAGCG 1320

GAAGTTGAGA GTAAACAAGC TGAGGCTACA AGGTTAAAAA AAATCAAGAC AGATCGTGAA 1380

GAAGCAGAAG AAGAAGCTAA ACGAAGAGCA GATGCTAAAG AGCAAGGTAA ACCAAAGGGG 1440

CGGGCAAAAC GAGGAGTTCC TGGAGAGCTA GCAACACCTG ATAAAAAAGA AAATGATGCG 1500

AAGTCTTCAG ATTCTAGCGT AGGTGAAGAA ACTCTTCCAA GCCCATCCCT GAAACCAGAA 1560

AAAAAGGTAG CAGAAGCTGA GAAGAAGGTT GAAGAAGCTA AGAAAAAAGC CGAGGATCAA 1620

AAAGAAGAAG ATCGCCGTAA CTACCCAACC AATACTTACA AAACGCTTGA ACTTGAAATT 1680

GCTGAGTCCG ATGTGGAAGT TAAAAAAGCG GAGCTTGAAC TAGTAAAAGA GGAAGCTAAG 1740

GAACCTCGAA ACGAGGAAAA AGTTAAGCAA GCAAAAGCGG AAGTTGAGAG TAAAAAAGCT 1800

GAGGCTACAA GGTTAGAAAA AATCAAGACA GATCGTAAAA AAGCAGAAGA AGAAGCTAAA 1860

CGAAAAGCAG CAGAAGAAGA TAAAGTTAAA GAAAAACCAG CTGAACAACC ACAACCAGCG 1920

CCGGCTCCAA AAGCAGAAAA ACCAGCTCCA GCTCCAAAAC CAGAGAATCC AGCTGAACAA 1980

CCAAAAGCAG AAAAACCAGC TGATCAACAA GCTGAAGAAG ACTATGCTCG TAGATCAGAA 2040

GAAGAATATA ATCGCTTGAC TCAACAGCAA CCGCCAAAAA CTGAAAAACC AGCACAACCA 2100

TCTACTCCAA AAACAGGCTG GAAACAAGAA AACGGTATGT GGTACTTCTA CAATACTGAT 2160

GGTTCAATGG CGACAGGATG GCTCCAAAAC AATGGCTCAT GGTACTACCT CAACAGCAAT 2220

GGCGCTATGG CGACAGGATG GCTCCAAAAC AATGGTTCAT GGTACTATCT AAACGCTAAT 2280

GGTTCAATGG CAACAGGATG GCTCCAAAAC AATGGTTCAT GGTACTACCT AAACGCTAAT 2340

GGTTCAATGG CGACAGGATG GCTCCAATAC AATGGCTCAT GGTACTACCT AAACGCTAAT 2400

GGTTCAATGG CGACAGGATG GCTCCAATAC AATGGCTCAT GGTACTACCT AAACGCTAAT 2460

GGTGATATGG CGACAGGTTG GGTGAAAGAT GGAGATACCT GGTACTATCT TGAAGCATCA 2520

GTCCCAGTTT CACTAGCTTT TCCGACCATA CGAATGTTGA GAAGGCCAAC GACCGTACCG 2940

ATAAGCTTGC TCAAACGGCC GTTCTTCACC AAGTTATCGA CTTTGGCTAG GACAAAGAGC 3000

AACTTAGTTT TTTCTTGATA GGCGGTGATA GCTTCAACCA CTTCTTCAAA AGACAAGCCC 3060

TGGTCAATCA AGTCATTCAA TTTTTCTACG AGTAGGTCAA CTTCACCACC AGCAGATAAA 3120

CTATCAATCA CATGAATCTT AGTGTCAGGA TGGTCTTCCA GATAAATATT CTTTGCTAGT 3180

TGAGCACTAT TGTGACTGCC AGAAAGGGTA CCTGTGATGG TTACTAGGAA AATGTTTTTG 3240

GCACCTTCAA ATGCTCGCAA ATAGTCATCT GGGCTTGGAC AAGCCGATTT TGAAGCTTCT 3300

GCAGTTGCAT ACATGGTTTC CATCATTTGG TCAATATCGA GACTGGCGTC ATCAACAAAG 3360

ACCTGATCAG CTACTTGAAT GGTTAAGGGG ACACTTACAA AGGTTGTGTT AATAGCTGGT 3420

GTTGGCAGTT GACGATAATC ACAACCAGAG TCAGCAATAA TCTTCCAAGT CATAGAAATT 3480

CTCCATCTTT GTCAGGAACG AT 3502

(2) INFORMATION FOR SEQ ID NO: 134:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12665 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

CGATTGATTT TTTTAAAGCG TTCGATAGAG AATGAGAAAC GAATCCTTAG CAATGGCGGG 60

AAAGAATTTG GAGTTGAGAA TACAAAACGA TTAACTATGG CTCATATTGT TTTTTATCTC 120

TCTTGCTTGG TTGAGGCAAT GGTGCACAAG ACAATTTTTG ATGGCATGGG CATGGTTGGT 180

TTAGTCTTGC TTATTTTTTC TATGCTGATG TTGATGTTGG TGATTCACTT GTTGGGAGAT 240

ATTTGGACAG TGAAGCTTAT GCTTGTCAAT AATCACAAAT ATGTAGATCA TATCTTGTTT 300

AGGACAGTAA AACACCCTAA TTACTTTTTA AATATTCTTC CTGAGTTGAT TGGCTTGACC 360

TTGTTGAGTC ATGCTTATGT GACTTTTGTT TTAGTTTTTC CAGTTTATGC AGTTATTTTG 420

TATCGACGAA TAGCTGAAGA GGAAAAGCTA TTACATGAAG TTATAATCCC AAATGGAAGC 480

ATAAAGAGAT AAATACAAAA TTCGATTTAT ATACAGTTCA TATTGAAGTG ATATAGTAAG 540

GTTAAAGAAA AAATATAGAA GGAAATAAAC ATGTTTGCAT CAAAAAGCGA AAGAAAAGTA 600

CATTATTCAA TTCGTAAATT TAGTGTTGGA GTAGCTAGTG TAGTTGTTGC CAGTCTTGTT 660

ATGGGAAGTG TGGTTCATGC GACAGAGAAC GAGGGAGCTA CCCAAGTACC CACTTCTTCT 720

AATAGGGCAA ATGAAAGTCA GGCAGAACAA GGAGAACAAC CTAAAAAACT CGATTCAGAA 780

CTTGTCGTCA TACCAGTTTT TGATACCACA GAGTTGGGGT TTGGAAATGA TCAGACCAGT 1200

TTCTTCTTTC ACTTCACGAA TGACAGCATC GACAAAGGAT TCGCCACGTT CAACATGACC 1260

ACCAGGAAAA GTAATGCCAG ACCAGTCGGG ATTAACTCGG TCTTGGACCA GGACCTTATC 1320

TCCGTTTTTA ATCATACACA TGTTAACAAA TTCGACTGCC TCTCTTCTGT TCATTCTTCA 1380

CAACCTTTAA TCTTTAATCA TAATGCAGAC TTCCCGCCAC CCAGCCGGTA CAGAGGGCAG 1440

AAGTGATGTT AAAGCCACCC GTGTGGGCAT TGATATCCAT AACTTCGCCT GCAAAGTGGA 1500

GGCCAGGTAC CAGCTTACTT TCAAGGGTTT TAGGATTGAT TTCCTTGAGA CTGACTCCAC 1560

CCTTGGTAAC AAAGGACTTT GCAAGGGACA TTTTTCCAGT TACAGGAATT TTAAGTTCTT 1620

TAATGGACTG GACAAGTTGT TCTCGTTCCT TTTCAGTCAG TTGTTTGACT TTTTCAGGAT 1680

ATCCTTGTAC AAAAAATTCG GCCAAGCGTT CTGGTAACAA GGTTTTTAAA GCGTTTTTCA 1740

AGGATTTTTC CCGATTTTCT TCTAGAAATG TAACCAAGTC CTTCTCAGAA AGTTGAGGCA 1800

AAACATCGAG TGAGAGAACC TCCCCACCTT TGACAAAGCT AGACATGCGT AGGGCAGCAG 1860

GACCTGACAA ACCAAAGTGG GTAAAGAGTA AATCATGAGT GATGACATGC TTACCATAAC 1920

TTAGGGTCAC ATCGTCCAGA GAAATACCTT GTAAGGCTTT ATGTGGAAAA TCTGTTAATA 1980

AAGGACTTTC AGCAGCCTCA AGATCGGTGA TGGTATGCTT AAAATGGCGA GCAATCTCGT 2040

GACCAAAACC AGTCGAACCA GTCGAAGGAT AAGACTTACC ACCTGTTGTG ACAATGAGTT 2100

TCTCACAAGT GAAGGTTTGA TCCGCTGACT TAAGGACAAA CTGGTCATCT ACTTTTTTAA 2160

CAGAAACGAT TTCTATTTGA GTAGCAACTT GACCACCTAG TTCGGTGATT TTCTTTTCCA 2220

AAGCTTCGAT AATAGTCCGA GACTTGTCAC TGGCTGGAAA GACGCGTCCG TGGTCTTCGA 2280

CCTTAAGTTT AACACCATTT TCTGTAAAAA AGTTGATGAT GTCATGATTA TCGAACTGGG 2340

AGAAAACACT GTAAAGAAAG CGTCCGTTTC CAGGAATTCC AGCTAGCAGG TTGTCTAAGC 2400

TACCATTGTT GGTCACATTG CAACGTCCCC CACCAGTCCC AGCTAATTTT TTTCCAAGTT 2460

TCCGATTTTT TTCGATGAGG AGGGTTTTCT GTCCATAAAA GCTACTGGAA ATCGTAGCCA 2520

TCATACCAGC AGGTCCCCCA CCGATGACAA TAGTATCAAA ATGTTTCATA GCTCTATTGT 2580

ACCACAAAAA AACAAGAGAT GATGGTCACC TCTTGTCAAG AATGCAATTA ATCAATTTCA 2640

TAGCCCATCA GCAAACCGCC CTCTTCTGCA TAGAAACTGC AGAGACCAGA GGTTGGTAGA 2700

ATTTTAATAT CCGCTTGTGG GAAGGTTTCA CGGATTCGCT CTGAGAGCTG TTGACAACAT 2760

TTTTCGTTAT TGCGTTGGGC CATGACAATA CGGCCACCAG CATATCCAGC TTTTACTAAC 2820

TCATCATAGG CAGCTTGAAC TGATTTCTTT GATCCCCTTG CTTTTTGTAG CAATTCGAGA 2880

TTTTAACTCT ATTTATTACT AGATTTCATA ATTAAAAAAC CTACTGACCA AGCTAGAAAG 9360

CTTGATACAA TAGGCTTTTT AAAGACTGAT TATTTAACAG CGTCTTTAAG AGCTTTACCA 9420

GCTTTGAATG CTGGTACTTT AGAAGCTGCA ATTGTCATTT CTTTACCAGT TTGTGGGTTG 9480

CGACCTTTAC GTTCTGCGCG CTCACGAACT TCAAAGTTAC CAAAACCGAT CAATTGAACT 9540

T 9541

(2) INFORMATION FOR SEQ ID NO: 133:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3502 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

TTGACTATCC TATCATGCTT TCTAAGGTCT ACTCAAGAAA ATCATTTTCA AGTTTTCACA 60

CCTTTCTCAA AAAAGTTAAA AAATTTTCTC AAAAACGCTT GACTCTGACC TAAGGCGAAG 120

GGTTATACTA TCATTGTAAG GAGGAAATCA TGTACCATAT AAAAGAAGCT GCGCAGCTTT 180

CAGGTGTCTC TGTCAAGACC CTGCATCACT ATGACAAGAT AGGACTCTTG GTCCCCTTAA 240

AGTCGGAAAA CGGCTATCGA ACCTACAGTC AAGAGGATTT GGAACGCCTT CAGGTCATTC 300

TTTACTACAA ATATCTAGGC TTTTCTTTAG AGAAAATAGC AGAGCTGTTA AAGGAAGAAA 360

GGACAGATTT ATTGCCCCAT TTGACTAGGC AGTTGGACTA TCTAACTCGC GAAAGGCAAC 420

ATCTGGATAC CTTGATTTCC ACCTTGCAAA AAACTATTCA AGAACAAAAA GGAGAAAGAA 480

AAATGACCAT TGAGGAAAAA TTCACGGGAT TTAGCTATCA AGACAATCAA AAATACCACC 540

AAGAAGCGGT AGAGAAATAT GGTCAAGAAG TCATGGGACA AGCGCTCGAA CGCCAAAAAG 600

GTCACGAAGA CGAGGCTACG GCCGCCTTTA ACCAAGTCTT TCAAACTTTG GCACAAAATC 660

TTCAAGTTGG TTTACCTGCA ACAGCAACCG AAAACCAGGA GCAAGCAGCC AAGCTCTTGC 720

AAGCCATTCG CACTTATGGA TTTGACTGCT CTATTGAGGT ATTCGGTCAT ATCGGTAAAG 780

GTTACGTCTA CAACCCAGAG TTTAAGGAAA ACATTGACAA GTTTGGTTCT GAAACAGCCC 840

AGTACACGTC AGATGCCATT GCGGTTTACG TTCAGACAAA TGCAGAATAA ATAGGCTAGG 900

AATTTCCTAG CCTATTTTTT ACTTCAAATC ATAAAGCCAG TCGTCACCGT TTTTGTAGTA 960

AAAGAATTCA CTGAGATCTT CTTCTAGAAA CACACGAAGC ATATCAGACA TATCATCGGT 1020

TGCAAGTTTT AGATGAGAAA GATTTTCAAA GTCCTCCCAC CAAACTTTCC CTTCGTCTGA 1080

AGACTGGAGT TCACCAGTAA AGTGTTCTGT CTTGTAAAAA AGGACGACAT AACGATAATC 1140

ATCCAGAATC AAAAAGTTAT TGTTTTCCAT AGACAATTTA GCTAAAAGCA AACGAGCTTT 7620

TTCGCCACCA GATAGCATGC CGACTGATTT TTTAACATCA TCTCCTGAGA AAAGGAAGGC 7680

TCCAAGACGG TTGCGGATTT CAACTTCTGG TGTCAGTTTG AAATCATTCC AGAGTTCATC 7740

CAGCACCGTA TTACTTGGTG TCAGCTTGCT TTGGGTTTGG TCATAGTAAC CAACCTCAAC 7800

ATTAGCGCCA AAGCGCTTTT CTCCCTTGAT AAAAGGAATC TGGTCCACAA TAGACTTGAT 7860

AAAGGTTGAC TTGCCGATAC CATTTGGACC AACGATAGCG ACAGCATTCA TCTTACGAAG 7920

ATCTAGGTTA ATCGGTTGTG ACAAGACTTC CCCGTCATAG CCAACAGCTG CATTTTCAAC 7980

AGTCAAAACA ACATTGCCCG ACGTTTTTTC AGACTGGAAG GTCATGTTGG CTGATTTCTT 8040

GCCAGCTTCA GGCTTGTCCA AACGTTCCAT TTTTTCCAGT TGTTTACGGC GAGATTGAGC 8100

ACGTTTAGTC GTTGAAGCAC GAACTAGATT GCGATTGACA AAGTCTTCCA GAGCAGCGAT 8160

TTCCTTCTGT TGCTTTTCAT AGTTTTTTGC CTCAGTAACT AGCTTTTGCT CCTTCAATTC 8220

GACAAAACGA GAGTAATTCC CCACATAGCG ATCCAAGGAA TGCTTGGTCA AATCTAGCGT 8280

AATTGTCGCA ACCTTGTCCA AGAAATAACG GTCGTGGCTG ACGATAATGA GGGCACCGCT 8340

ATAGTTTACC AAGTAATTCT CTAGCCAGGC GATGGTTTCA ATATCCAAGT GGTTAGTTGG 8400

CTCGTCCAAG ACCAAGAGAT TGGGCTTTTC AAGGAGCATT TTGGCAAGTG CCAAACGAGT 8460

ATTTTGACCA CCAGAAAGCT CAGCAATTTT CATCTGCCAC ATAGACTCGT CAAACTTGAA 8520

TCCATTCAAA ATCGCTCGAA TATCAGCTTC ATAGGTAAAG CCACCTGCTT GGCGAAAATT 8580

CTCAGATAAG CGGTCATAAT CTGACATCAG TTTATCCAAA TCCTCACCAG ACTTTTCACC 8640

CATCTCCAGC TCCATCTGAC GCAGTTGTCT CTCCGTCCGA CGCAAATCAT TAAAGACATG 8700

AAGCATTTCA TCGTAGATGG TATTTTCAGA CTCAAAACGG CTATCTTGGG CTAGGTAAGA 8760

CAGAGAAATA TCTTTTTTCT TATTGATTTC TCCGCTAGTT GGCTCCTCTT CTCCAACTAA 8820

AATCTTCAAA AGAGTAGACT TACCTGCACC ATTTTTCCCA ACAAGAGCAA TCCGATCTCG 8880

TTCATCAACC TGCAGGTTGA TATTATCGAA AAGAACCTCT CCTGCAAAAG AACGTTCAAT 8940

TTTATTAGCT TGTAAAATAA TCATACAAGT AGTATAGCAT GTTTCCCTAA GGCATTCAAG 9000

ATAATCGTAA GTCTTTTAGT ACAACTTTTA TAACATAAAA TAAACTAAAT TATGTATATT 9060

TTATATTAGA TTACTTCACT ATCTTGTTGG ATTTTCTAAC CAGCTAATCT TGTTTCAAAT 9120

AGTTATCGCA CAAGTCTATT ATTTAATTCT TTTCATCATT TACGTACGTA TAGCAGATTG 9180

AAATAAGATG AGAACAAATC GATTGGGAAA GTAAAATTAA TTTCTATAAA TGTTTTAGCA 9240

ATTGTTTCGT ACTATTTTAG ATTCAGTCTA CTATATACAA TATTTTCGGA ACATTCAACT 9300

AGCTTTATTG GCTGTGCGAG AAAAGGCAAA AGCAATAGCC TGCACAATGT TAGCAACAAT 5820

CAAAGCGCCA ATCATCCAGC TATCATTCCT TATGAAAGAA ATAGCCAGAC AAAGAATCCC 5880

ACAAACAAGA TCTGCCGTCA TTAAAATCTT ACGACGAGAA AAACGGTCTG AAATAACTCC 5940

GCCAAAGGGA TTGACGAGAA TAGATGTGAC GAGCTCAGAA ATCTGATACA TTCCTAAAAC 6000

TGTCTGTCCT ATAGTCCCCA TAGAAGCCAA CCAGACACTA TTTCCATAAT CATAGAGCAT 6060

ATTTCCCATT TTATTGATAG CCCCACGGCT AATCAACTGC ACTGCATAGC GATTCATATT 6120

AAAGCTCCTC TCAAATTTTG AAACTATTGT ATCAAAACCG AAAGGAGCTT TTTATTTTTT 6180

CCCTTATTTG GGAAAATTAA CTTTTGACAA ATTTTTCGTA GTGTTCCTGA TAATAGGCTA 6240

CTTGCTCTGG AAGACCTAAC ACATCAAAAA TATGCATGGC CTCTTGCATC TGCTTACAGC 6300

CTTCTTTACA CTGTCCTTTT TGATATAAGG CAAAACCTTT TAAATAATGG AAAACATTAC 6360

GCTCATAAAG CTTAATACCT TTGTCAATAA TCTTCTCTGT ATAAGCCTCA AAATAGTTGG 6420

CATTATAAAA AGAAGAATGC TCTAAACAAT GCTGGTAACA ATTGAGGGCC AAAATCAACA 6480

CTAATCTCTT ATGGCGACTA ATCTCTTGGT AAAATTCCTC CCTCTCCATA ACTTCTCTAC 6540

CAATCCGAGT GACATAGTCT ACATCGTAGA AACTATAGAG GTTACCGAAA AGAATCAACT 6600

CATACATGGT CCATTCTTCT GTTTTGAAGA GATAATCTGC TACCTTACCC AAATCATCCT 6660

GCTTCATATC ATAACTCGCA TCTCTTTGAC AAATCAGACC TTGTAGCAAA ATCCAGTTCA 6720

GCTCAAAATA AAGGGGAGTC GTCGAACTCT TAGACTTTTC AAGTTGTTCT CTTTGAAGCT 6780

TTTGAAAACC TGCAATATCG TTTGAATAGT AAAGTGGGAT AATCTGTGCC ATCATAGACA 6840

CATGTTCATG ATTATGAAAA TTCCTTGCCT TATCCATGAA ATTTTCGATT GTTACATGAA 6900

TGTTATCCAA AATCTCAAAG AAACGGGAGA CTGCCAGGTC AGACTCCCCA AGCTCAAAGC 6960

GAGATAACTG AGAGGTAGAG CAGGATTCGC CTGCTGCTTC CTTTAAAGAA TAATTTCCAC 7020

TTGTTCGAAA TTCACGAAAT ACTTTTCCAA GATGTTCCAT CTTTACACCT GCTCTGATAA 7080

TTCTTCCCAC TCAAGCATAG CTTCTTCCTG ACGATGGCTG ATTTTGTCCA GCTCAGCCTG 7140

TAATTCCATG AGTTTGTCGG CATCGTTTGT TTCCAACATT TGTTCAGAAA TGGCTTGGCT 7200

TTGACTTTCT AGCTCTTCAA TTTCAGCTTC TAGACTTTCG ATTTGTCGCA TGAGTTTGCG 7260

AACTTCTTTT TGACTTTCTT TCTGGGCCTG ATAGTCATTG ACTGGACTTG CTTCCTTTGC 7320

TTGATTGCTA GTTGAAGCTT CCTCAGTCTG ACTCATTTCT GCTGTTGCTT TCTTCTCAAC 7380

ATAGTAGTCG TAATCTCCAA GGTAGAGAGT TGAACCATTC TCAGACAATT CCAAAACATG 7440

AGTTGCCACA CGATTGATAA AGTAACGATC ATGACTGACA AACAGCAAGG TTCCATCAAA 7500

GTCAATCAAG GCATTTTCTA GCACTTCCTT ACTATCAATA TCCAAGTGGT TGGTCGGCTC 7560

CCACTCCATC AATATCGTAA GGCAGATTTT CCCGTTCCTG TCCTACTTCT TGGATAAAAT 4080

TCCAGATTTC ATCTATGTTT TCAGCCAAGA TTCGCTTAGG ATTGACCACA AAACCTAGTT 4140

GTTCTAGGTA CTTCAAACCC TTTTCTTGGC TATCACGAGT TGAAGGGCTG GCTTCTTGAT 4200

AGAGAAACGT TGCAAGATTA CGCTTGGCAA CTACTGCTGT ATCCAACTGA CGCAGAGTTC 4260

CTGCTGCCGC ATTACGAGGA TTAGCAAATT CAGGCTCTCC ATTTTCTTGG CGCGCTTGGT 4320

TAACTTGGTC AAAGGAAGCG CGTGGCATGT AACATTCCCC ACGAACTGTG ATATCTAGTT 4380

CTTCTGGCAA AGTCAAAGGG ATGTCCTTAA CACGCTTGAG GTTTTCTGTG ATATTTTCAC 4440

CAATTGAACC ATCTCCACGT GTTACCCCAG CAACCAAAAT CCCCTTTTCA TAAGTCAGCG 4500

AGATAGATAA GCCATCGATT TTCAGCTCAC AAATATAGGT CGGATGAGCC ACTTCCTTAC 4560

GAACACGCGC ATCAAAAGCA TCTAGCTCCT CACATGAAAA AGCATCCTGC AAACTATAAA 4620

GAGGATACTG ATGACTGTAT TTTTCAAAAC CATCTAAAAC CTTGCCACCA ACACGATGAG 4680

TCGGACTGTC TGCTAGCACT TGCTCTGGAT AAGCAGTTTC TAACTCGACC AACTCACGGT 4740

AAAGGCGGTC ATACTCACTG TCTGAAACCG AGGGATTATC GCTGGTATAG TACTCAGTCG 4800

CATAGCGATT GAGCAAAGCG ACTAACTCAT TCATTCTTTT ATTCATAAGA CCATTTTACC 4860

ATAAAACAAG CCCTCCTCAC AAACGAGAAG GGCGGAAAAA ACACTTAGTT TGAAATTATT 4920

TTTGAAACTC AAGCAACCTT ATATCAATTT TTCAAAATGA GTTCGAACAT ATCCGAGAGC 4980

TAAGAAATAT AAGGCTACAA CTCCAAGTCC AATAATCAAG AAAGAATAAA GATGGACACT 5040

TGGCAAGACT GTCATAAATC CTTTTGCAAT AGGCATAAAT AGAATAGCTA AGGTAAAAAT 5100

TGTACTCAGT ACTCTTCCAA GAAATTCGCT CTCAACCTTG GTTTGTACTT GAGTAAAAAA 5160

GTGAATATTA AAAATCGTCA TAAACAATTC ACAAACTAAA TTTCCAGAAA AGGAAAGAAA 5220

AGTTGGAAGT GGTAATCCCA TCATAAAAAC TCCGACACCT GTCAAAGCCA GTAAAATCAA 5280

AAGATTATAA ATATTAGCTT TAATTTTACT AGCTAGAAGA GCCCCAATGA TGGAACCAAT 5340

AGCCCCCATA GTTAAAATAC TTGCATAGGC TCCTTCTGAC CCGTAAAGCT GATTCGAAAA 5400

GGGAAGTAGA AATTCAAAAG CTGCAAAAAA GAAATTAACG CTGGAAGCTA CCAGCAAAAG 5460

GAAGAAAATT TCTTGCTGAT GCCAGATATA GTGTAACCCA TCCTTGATAT CTACAAAAAT 5520

ATCTCTCCCA GTAAAAGCCT TTTTCTCTTG AACTTTTGCT TCCTCTTTTG GAAGGAAAGC 5580

CACTAGAACA AAAGCAATGA AAAAAGTCAG CGAGTCTAGC AGTAGCGTCA TATGGAGACT 5640

TGCAAACTGT AAAACAAGGA AGGAAAGAAC AGGAGAGCTA ACACCTACAA CCTGCAAAAC 5700

CAGCTCTAAG CGAGAATTAT AGATCACAAT CTCATCTTTC TCCACCACTT CAGTTATGAT 5760

TGTCCTTATG ATTATAAGAG TAGCTATTGC CTCTTTTCAT CTCAAAAGTC TTCCAAACGG 2280

GTGCATCATT AGCAGCTGAT TCATAAACGA CAACTTGCAC TTCTGTCGCT GTAGGTGACC 2340

AGAGAGAAAA ATGAGCCTGA TTGTCCTCTA CACGGCAACC CAATTCTCCT TGGTAACCCC 2400

AATGATGATC AAAACTAGCA CTGTTAATGG CCTTATCAAA GGCAAAAGGA TTTTGATTTT 2460

TATAGAAAGG ACTGGCAATA GCAGGATTTT CAGAGTAATA AATCCTATCA TCGCCTTCCA 2520

AAATCCAGAC CTCTGTTAAT AGGGGATAGT GATTAAAACG GATAGAATAT TCTTTACTAG 2580

TTTGACCTGT ATGAACCACA AAATTCAAGC TTTCTATAAC ATGTGAACTT GGGTGTTCAA 2640

AGCTAAATAA AGCTCCAAAA TAATCTTCTT TGTAGGTTAG CAAATCAATT CGTTGATCCT 2700

GACTTTTTAC AAAGGAGCAA GTGTCATATT CTCCATTCTT ACGATGGTAA TGAATGCGCA 2760

TAGGGTAGTT ATACATTTTT TATTTTTCCT TTTTACTTTG TTTCTATTTC ACTAATAAAT 2820

TTTTGTCAAT CTCGTCTCAA TTAACAGACA TAGTCATATT CTCTAAACTC TGTTTTTAAA 2880

CGATCCATTA CAAACTTTCT AGCCATGCCT CATCTCTGAC CTGGATACCA AGTTCTTGTG 2940

CTTTTTGCAG TTTACTTCCA GCGTCTGCAC CTACCACGAC GAGGTCGGTC TTTTTAGAAA 3000

TACTACCTGT CACTTTGGCA CCCAGACTTT CGAGTTTACT TTTAGCTTCT GAGCGCTTGA 3060

GTCGTTCCAA TTTTCCTGTC AATACCACGG TCAAACCTGA CAAGGCCGCA TCCGCTACTA 3120

CCGTCTGTCC TTTATAGTCC AGATTGACCC CAGTTTCTTT CAATTCTCTG AGCAGAATTT 3180

CAGAGCCTTC TGTCGCAAAA TAAGTCTGAA GACTTTTGGC AATCACGCCA CCTAGACTTT 3240

CAATACTAGC CACTTCCTCT GAATCTGCCT GAGACAGATT TTCAATTGAA TGGAAATATT 3300

GAAGTAAAAG CTGACTAACC TTGCTTCCGA CATGACGAAT TCCCAAACCA AATAAGAGCT 3360

TCTCGGCAGA ATTTTCCTTT GATGCTTGGA TAGCCTGATA CAGTTTAGCA GCGGACTTTT 3420

CCTTAACTCC CTCTAAAAGG AGGAAATCCT CTTCTTGCAA ACGATAAATA TCCGCCACAT 3480

CCTTGACTAA ATTAGCAGCA AAAAGCTTCT CAACAATAGA TGGACCAAGG CCTGTAATAT 3540

TCATAGCATC ACGAGAAGCA AAGTGAATCA AGCCTTCCAT GATTTGAGCA GGGCAACGCG 3600

GATTGATACA ACGTAGGGCC ACTTCATCTT CAAAGTGCAA CAAGTCAGAG TTACAACTTG 3660

GACAGTTTGT AGGGATATCT AGTTTTTCTT CAGAAACCCG TTTGGACTCT ACCACACGTA 3720

AAACGGCAGG GATGATGTCA CCAGCCTTAT ATACAATGAC CGTATCGTCT TTTCGGATAT 3780

CTTTTTCAGC AATATAATCT ACATTGTGCA GGGTCGCACG GCTAACAGTC GTACCGGCAA 3840

GTTGTACTGG TGTTAGATTA GCAGTTGGAG TTACAACACC GGTACGGCCA ACTGTCCAGT 3900

CAACTGATAA GAGTTGAGCT TCTTTTTCTT CGGCAGGGAA CTTGTAGGCT ACTGCCCACT 3960

TTGGAGCCTT AACTGTAAAA CCAAGTTCTT CTTGACTTGC TAGGTCGTTG ACCTTGATTA 4020

GCGAGATAGA TGGTATTTAT AAAACACTCA AGACAGCTAG ACTAATATCA TTTAAAACAT 540

TATCTTCTTT TGAGCGACTG TTGGTTACCA ACATAGCTAA ATTTCCTGCA TTTTCAAATT 600

GATAGGGTTC TGATTTAGCA TTCACAACCA CCAAGAGGTG TTCTTTGCCG TGAACTTCAT 660

AGATAAGGTA GCCGCTATGT TCAATCGCAG AATGCACAAA GACATGATGG TAAATTTCAT 720

CATAGCTAGA GTAAGAAAAG GCACCAGTTT TTGTCTTCAA TCGGATGACT TGACGGATAA 780

ACTCAATACT GTCTTGACGC TCATTAATCA AGTTCCAGTT CACTTGGTTC ACACTGTCAG 840

GAGCATTATA GCTATTCATC GCACGCTCTC TATCATCATG GGTCAACTCA CCATTTTCAC 900

CAGTCGCAAC CAGTTTGGTA CGACCAAATT CTTGACCGAT TTCCATAAAG GCCATCCCCT 960

GCATGAGCAG ATTCATGGCT GTGGCAGTTT CGACCTTGCG CATGATTTGC TCTGAACTTT 1020

GGTCTGGATG AAGGGTTGCC AATAAATCGT GAAGATTGTA ATTGTCATGG GCTTCTACAT 1080

AGTTAAGCAC CTGATTTGGA TGTGTATAGC TTCCTAATTC ACGACTTCCT AGGATTGCTT 1140

TAGCTAGAAT TGGCTCTGTC GCAGCACCAC TGACAAAACC TGACTTGATA GCACCATAAA 1200

CTTCTCCCCC TTTGACAGCA TCGCGCTGAT TGTCATTAAA GAAACCAATA TTTGGCATCT 1260

GGTAGGCATT GTCCTTCTTG GCCTTATCAT AAGGGGCAAG ACCTGTTCCC ATATCCCATC 1320

CTTCTCCATA GAGGATAATG TTGGAGTCGA TTTCATCCAA GCTTTGACGA ATCATCTGCA 1380

TGGTCTTGAC ATCATGAATC CCCATCAAGT CAAAACGGAA GCCGTCAATA TTATATTCCT 1440

GCACCCAGTA TAGAAGAGAA TCAATCATAT ACTTGCGAAA CATTTCGTGT TCACTGGCTG 1500

TTTCATTTCC AACACCCGTT CCATTCTGGA AGGTACCATC TGGATTCATA CGATAATAGT 1560

AATCAGGGAC TGTTGTTTGG AATGGTGCAT CAACAACTGA GAAGGTATGG TTATAGACTA 1620

CATCCATAAT GACTCCAATA CCCGCATCGT GATAAGCTTG AACCATCACC TTCAAATCAC 1680

GAATGACCTG AGCTGGATCA TCTGGATTAG TTGAAAAACT AGTTTCTGGC GCGTTATAGT 1740

TTTGTGGATC ATAACCCCAG TTGTAGGTTA CATTTCCATC CTCATCGTAT TCTTTATCAC 1800

GGTCTGCAAT TGGTTGCAAT TGAACATAAT TGTAGCCCAG CTTCTTGATG TAATCAAAAG 1860

CAGTTGACTG GCCGTATTGG TTAACTGTTC CAGCCTGAGC AGCACCCAAG AAAGTTCCTC 1920

GAAGATGTTC ATCTACACCC GATGTAGGTG ATTTAGTCAA ATCACGAATG TGCATTTCAC 1980

AGATAACTGC CTTACATGGA TTTTCCAAGC GCCAAGTAGC CTCCGAACCG TGCTTAACCT 2040

CGAAGTTTTC AACTTGCTTT TCTACATGGC TCAGAATAGC TGAACGTTTG CCATCAGGGC 2100

TGGTCGCGAT TGTATAAGGA TCACGTGTCA GTGTTTGGTG ATGAGGGAAT TGGACTTGAT 2160

ACTGATAAGT CTTACCTACC AAATCTTCTT CAACATCCAA ACTCCAGACA CCGATTGTAT 2220

TACCCCAACG CCCCAAACTT TGAGCAATCA TAACGCTAGG CGCCGCAATA TCTAGAAAAT 5340

CCCAAGTATT GATGAGTTTA CGGTCAGCAA AGATATAGAG CACAAGAGCC CCAGTTATCA 5400

AACCACCGTA AATGGCCAAA CCACCATTCC AAATGGCAAA AATCTCTCCT AAATTCTGAC 5460

TATAGTAATC AAATCGGAAA ATAACATAGT AGAGACGAGC TCCTAAAATA GCCAAGGGAA 5520

AGGCTACTAA GATAAAATCT AAAATATCGT CTGGTATGAT CTTCTTTCTA GGTGCTTCTT 5580

TCATGGTCAA ATAAACCGCA AGAATCAAGC CTGTCACAAT ACATAAGGCA TACCAACGAA 5640

TGGCTAGGGG TCCTAGTTGA ATAGCAATTG GATCAAGCAT TTTGCACCTC ATTTCGAGCG 5700

ATTAGACTTG TCAGTCGTTC GTCGAACAAA CGGGTCGCAT CAAAGCCCAT TTCCTTGGCA 5760

CGATAATTCA TGGCAGCTGC CTCAATCACA ACAGAGATAT TACGACCTGT TTTAACTGGA 5820

ATACGAATAC GAGGAATGtA CGCCAGAAAC TTCAAGTTCC TCTGCATTAT TTCCAAGACG 5880

ATCAAAGGTC TTATGCGTAT CGTAATTTTC CAAATAGACA GCAAGCTGAA CCTGTGAAGA 5940

ATCCTTGACA GCACTCGCAC CGTAGAGACT CATAACATCG ATAATACCAA CCCCACGAAT 6000

TTCAATCAAG TGTTTCAAAA TTTCAGCTGG TTCACCCCAG AGAGTAATCT CATCCTTGGC 6060

AAAGATATCG ACACGGTCAT CGGCTACCAA ACGGTGACCA CGTTTGACAA GCTCAAGACC 6120

TGTCTCGCTC TTACCAATTC CACTATCTCC CTGAATCAAG ACGCCCATCC CATAAATATC 6180

CATCAA 6186

(2) INFORMATION FOR SEQ ID NO: 132:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9541 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

GAAAATCACA ACCCTTTTTG CAAAATTTTT GAGATTATTT TCACAAACTT GATTTTTCAA 60

AGTATACTCA ATAAAAATTA AAAAAATCCA CTACGTCAAG GCGAGGCTAA TGTGGTTTGA 120

AGAAATTTTC GAAGAGCGTG AATGAGTATC ATCTATAGTA AAATAAAAAA ACTGAACAAT 180

TTGGTTGGGG ACAGCCAAAC CAATTTCTCA CAATGTTTCA GAAACAAGGG TGTGCTATTC 240

CAATTTCAGC CTACTATAAC TGTCATAGAT TGCTGAAACA AAGTCTAGGT AAAAGTCTTC 300

ATAATAAAAA GACCTCCTAT CAAGTGTTCA AAAACTTTGA TAGGAGGTCT TGTTTTGTGA 360

AAATATTTAT CAAATTTTCT ATACAAGTGA GCTGTTAGCC AGGTTCTTTC TATTCTTTCA 420

ATTTCAATGA ATGGATTTTT TACTAATACT CATAACTGGG AATTTGTCTG TGTAAAAATA 480

TAAATCAAAT CAATGGAGAT ATTGTCAAAA CCAGCCAGTT TCAGGCGATC GATATTTTCA 3600

TAAATATCCT TCTCCAAATG ACTGCGCCCA ATCTTTTTCA ACATCTTATC ATCAAAGGTC 3660

TGGACACCTA GCGAAACACG ATTGACAGCC GAATTTTTCA AAACAGCTAT CTTATCCGCA 3720

TCCAAATCGC CTGGATTGGC TTCAATGGTC AACTCTTCCA AGACAGACAA ATCCAAGTTT 3780

TTAGTCAAGC CATTCAGTAA CACCTCCAGT TGCGGAGCCG ACAGGGCTGT CGgTGTTCCA 3840

CCACCGATAT AAAGGGTTGA CAACTTTTCA ATATCATAAG AACGAAACTC TTCCAGCAGA 3900

TGCTCTAAAT AGCTGTCGAC TGGCTGATTT TTGATGAAGA CCTTTGAAAA ATCACAATAA 3960

TAACAAATCT GGGTACAAAA TGGGATGTGC ACATAGGCTG ACGTTGGTTT TTTCTGCATA 4020

GTAATTATTA TACCACAAAG ACTAGATTCC AGATAAAAAT CACCATCCCC AGATACATAG 4080

TCCGTCCGGA GATGGTGATG GTTTATTCTT CTGTTATATC AATCACAATC TCTTCTGAGT 4140

CATCAAGAGC TTCGGCTTTT TCTTGCCATT GCTCCTTGAG ATTATTTAAT TGATTTTTTG 4200

ATGCTTCTGT CGCTTGAAAA GCATAGGATT TAGTTTGAGC AAGTATACTG TCCACAGTGA 4260

TTTCACCTGA CTCAACCTGT TCTTTTGTTT TCAGAACAAA ATCTGTAGCC TGCTCCTTAA 4320

CTTCTGTCAG TTTTTCACAG ACTTGCTCCT TGGCATACTC CGGATCTTCT CTCAAATCAT 4380

CTAGAAAATC TTGAGCCTGA CTGCAAACTT GTTTGCCCTT ATCACTTGTT AAAAACAAGG 4440

CAAGAGCTGC ACCTGAAACG GTTCCTAAAA GGATTGAGGA TAATTTACCC ATAAGGATTC 4500

TCCTTTTTTA TTTTTTGAAA AATTTACTTG CAAGACGAAG AGCTGACAGA CTTGCACCAG 4560

TCTTGAGTGT TTTTGAACCA GCTGATGAAG CTTTCTTGCT CAAGACACGC GCATGGTCAT 4620

TGAGGTCTGA AACAGATAGA GATAAATCTG CAACAGCACT GAAGAGTGGA TCAATCGTAG 4680

CCACCTTGAC ATTGATATCA TCTGCCAAGA CATTGACCTT AGCCAACAAC TCATTGGTGT 4740

GATGCAAGGT CACATCCACA TCTGAAGTCA AGGTTTTAAT CGTCTTTTCT GTTTCATCGA 4800

TGACACGACC AAGCTTTTGT ACAGTAATGA TCAGATAGAC CAAAAAGACA ATCAAAGCTA 4860

GGGCAACAAG AATATATGCA ACTTCTAACA TTTAGTTTTC CTCCTCTGTA ATATAGTAAG 4920

GGGCCTTCTT TCGATTTTGA TAAATAACGA TCATTATACC GAGACCGATA AGGACAACTG 4980

ACAGCCATTG GGACACTCGA AAGCCGAAGA ACATGAGACT ATCTGTTCGC ATACCTTCGA 5040

TAACCATACG ACCGAAACCA TACCAAATCA AGTAAAAGGC CGTGATATGA CCTCGTCTGA 5100

GACTCTTCCA TTTCCGTCTA AAAATCAGAA TCAAGGCAAA GCCAAGCAGA TTCCATAGAG 5160

ACTCATAAAG GAAAGTCGGT TGACGGTAGC TCCCCTCAAT ATACATCTGG TCACGGATAA 5220

AGCCAGGTAG ATAATCCAGA TTATCCACTG TTGCACCATA AGCTTCTTGG TTAAAGAAAT 5280

TTTATTCATA GAAATCAACT TGGGTATCCA ACAATTTATC CCCATCATAA ACAAACTTGG 10260

CTGAAAAGAA GGGTTTATCC TCTAAAAGCC ACTCAACAAA GGTGTGGTCA CCTTCCCAAG 10320

TCGGCTTGCT CAAAACCTCA TCATAGGGAA CCCATTCTAG CGTCCCCTCA TTGCAGTCAA 10380

TCAAGTCGCC CTCAAACTCC GTCACCTTAA AAACATAGGT GTACCAGTCT AAATCTGGTG 10440

TAAATTCAGG AAAAGTGATG ACACCTTTTA GAACTGGCTT GGCTTTGAGC CCTGTTTCTT 10500

CAAGGATTTC ACGCGCCGCG CATTCCTGGG GCGTCTCTCC TCTCTCTAGC TTACCACCCA 10560

CACCAATCCA TTTCCCTTCA TGGACATCAT TGGGTTTCTT ATTACGATGG AGCATGAGCA 10620

GTTCTTTCCC ATTATCAATG TAGCAAATCG TCGCTAACTG AGGCATATTT TCTCCTTATC 10680

TAAGCCAATC GATTGGCTCT TGTCCTGTCT CTTTTAAGAA TGCATTGGCC TTGGAAAAGG 10740

GCTTGGAACC CCAAAATCCT CTATAAACCG ACAAAGGACT TGGATGGGCT GATTCGATAA 10800

TCAAGTGATG AGGATTGGTA ACTAATGCCT TCTTCTTACG TGCATAAGCT CCCCAGAGTA 10860

CAAAAACGAC TGGTCTATCT AGATGATTGA CCACCTGAAT CACAGCATCA GTAAAAGGCT 10920

CCCAGATTTG ACCAGCATGA CCATTGGCCT GTCCAGCAGG AACAGTCAAA CAAGCATTAA 10980

GAAGCAAGAC TCCTTGCTCA GCCCAAGCTG TCAAATCATG AGATTTCTTA ACTCCGATAT 11040

CATCTGACAA TTCTTTCAAG ATATTTTGCA AGGATGGTGG AGCTGGGATA GAGTCAGGTA 11100

CAGAAAAACT CAAGCCCTGC GCTTGACCTG GTCCGTGATA GGGGTCTTGC CCTAGAATTA 11160

CCACCTTAAC TTCTTCAAGC AGTGTTGTCA AGAGAGCCTG AAAAACCTTT TCCTTGGGTG 11220

GATAAATAAT CCCCTGAGAA TAGACCTGCT CCATAAACTG ATTGATTTTC CCGAAATAAC 11280

CCTCAGGTAA TTGCGCCTTA ATCAAAGCAT GCCAAGACGA GTGTTCCATA GCCGACTCGG 11340

(2) INFORMATION FOR SEQ ID NO: 148:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12127 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

AAAAAATAGA CTTGTTAGAC TATAAATGTA GTAAGCCTAC ACAAGAAAAA TACATAGAGA 60

TAAAGGTGAT TATTATGAAA TTCAAAAAAA TGCTTACTCT TGCAGCCATT GGCTTATCAG 120

GATTTGGGCT TGTTGCCTGT GGCAATCAGT CAGCTGCTTC CAAACAGTCA GCTTCAGGAA 180

CGATTGAGGT GATTTCACGA GAAAATGGCT CTGGGACACG GGGTGCCTTC ACAGAAATCA 240

CAGGGATTCT CAAAAAAGAC GGTGATAAAA AAATTGACAA CACTGCCAAA ACAGCTGTGA 300

TTCAAAATAG TACAGAAGGT GTTCTCTCAG CAGTTCAAGG GAATGCTAAT GCTATCGGCT 360

ACATCTCCTT GGGATCTTTA ACGAAATCTG TCAAGGCTTT AGAGATTGAT GGTGTCAAGG 420

CTAGTCGAGA CACAGTTTTA GATGGTGAAT ACCCTCTTCA ACGTCCCTTC AACATTGTTT 480

GGTCTTCTAA TCTTTCCAAG CTAGGTCAAG ATTTTATCAG CTTTATCCAC TCCAAACAAG 540

GTCAACAAGT GGTCACAGAT AATAAATTTA TTGAAGCTAA AACCGAAACC ACGGAATATA 600

CAAGCCAACA CTTATCAGGC AAGTTGTCTG TTGTAGGTTC CACTTCAGTA TCTTCTTTAA 660

TGGAAAAATT AGCAGAAGCT TATAAAAAAG AAAATCCAGA AGTTACGATT GATATTACCT 720

CTAATGGGTC TTCAGCAGGT ATTACCGCTG TTAAGGAGAA AACCGCTGAT ATTGGTATGG 780

TTTCTAGGGA ATTAACTCCT GAAGAAGGTA AGAGTCTCAC CCATGATGCT ATTGCTTTAG 840

ACGGTATTGC TGTTGTGGTC AATAATGACA ATAAGGCAAG CCAAGTCAGT ATGGCTGAAC 900

TTGCAGACGT TTTTAGTGGC AAATTAACCA CCTGGGACAA GATTAAATAA AATGTTTGCT 960

CCATAAATCT CTAAAGAGAT GCAGACGTTT CATCGTACAA TAAGATAAAG AAGGCAAGTA 1020

GGGAGGTGTC GTATCTCCCT TACTTTCTTC ACTAGAAAGG ACAAGATGTG ACAAAACAAG 1080

CCTTCAAAGA AGCAGTTTTT AGGGCAATTT TTTTCATGAG TGCAACAGTA GCTGTTGTAG 1140

CTATTTTGCT AATCTGTTTC TTTATTTTTA GTAATGGCTT ACCTTTCATA GCTAACTACG 1200

GCTTTGCCCG TTTTTTATTA GGCAGTGATT GGTCGCCAAC GAACATTCCG GCAAGCTATG 1260

GTATTTTACC AATGATCGTT GGTTCCTTAT TAATTACCTT AGGAGCGATT GTGATTGGGG 1320

TGCCAACAGG CATCTTGACA TCGGTGTTTA TGGTTTATTA TTGTCCAAAG CCCGTCTATG 1380

GCTTCTTAAA ATCAGCTATC AACTTGATGG CAGCCATTCC ATCTATTGTT TATGGTTTTT 1440

TCGGCCTACA ATTATTGGTG CCTTGGATTA GAAGCTTTTT AGGAAATGGC ATGAGTGTCC 1500

TAACCGCTTC GTTACTATTA GGAATAATGA TTTTGCCAAC CATTATCAGT TTGTCAGAAT 1560

CTGCTATCCG AACAGTTCCC AAAACGTATT ATTCTGGTAG CTTGGCTCTA GGAGCTAGTC 1620

ATGAACGGAG TATTTTTAGT GTCATCTTGC CAGCTGCGAG ATCTGGTATT TTATCAGCAG 1680

TTATTTTAGG AATCGGTCGC GCAGTAGGTG AAACCATGGC AGTTATTTTG GTGGCAGGCA 1740

ACCAGCCGAT TATTCCAAGT GGACTCTTTT CAGGAACCAG AACCTTAACA ACCAATATTG 1800

TTCTGGAAAT GGCTTACGCA TCAGGTCAGC ATAGGGAAGC CCTTATTGCA ACCTCAGCAG 1860

TTCTCTTTTT CCTTATTCTC TTGATTAATG CCTACTTTGC CTACTTGAAA GGAAAATCAT 1920

CTTATGAGTA AATACCTGCT AAAACTTCTC GTTTATTGTT TTTCAGCTTT AACCTTTGGC 1980

TCTCTCTTTT TAATCATTGG TTTTATCCTC ATCAAAGGCT TACCTCATCT AAGTCTATCC 2040

GAAAACACCT TCATTGGTCG CACCATCTCC AAAGAAGCAG ACAACGATTT TACCGGTATT 6720

TTGCATTTGC TGACTGAGGG CTGCACCGAC AGCGATCCCC ATACCACCAC CTACGATACC 6780

ATTGGCACCA AGGTTCCCAG CATCAAGGTC AGCGATATGC ATAGATCCAC CTTTCCCTTT 6840

ACAGGTTCCA GTGTATTTAC CAAGGATTTC AGCCATCATT CCGTTGAGGT CAATCCCTTT 6900

AGCAATAGCT TGCCCGTGTC CACGGTGGTT TGAGGTAATC AGATCATCTG GATTGAGAGC 6960

TAACATAGCC CCCACGTTAG CTGCCTCTTC ACCAACAGAA AAGTGCGTCA TTCCTGGCAC 7020

TTTCCCTTTC TTTACTAATT GTGCAATTTT TAAGTCCATG CGACGGATTT CTTCCATCTT 7080

ACGGAACATT TCTAGCAAAA GATTTTTATC TAAAGTTGAC ATCTTCTTGC CTTTCTAACT 7140

TTCTTCTTAC CTTACTATTT TACCGCTTTT GGCAAATACT GTCAAAGTTT TTCTAAAAGA 7200

AATTTCACAA AATAAAAAAG AAAACCCCGT GAAAACAAGG GATTTTCTTG TCAAGAATAT 7260

TTTTTCACAA ACTTTTTAGC ATTTGGATTT TGCTAAAGAT TCAAATCTCT TCATAATCAC 7320

AGTTAAACGC CAACGGTAGA GCGCCCCGCT CACAATCAAA CTAATAATCA AGCCGATCCA 7380

GTAAGAATAA GCTCCAAAAT CTGTTAGGGA ATCAAATAGC GTAnCACAGG GATTGCTACG 7440

CCCCAATAAC CAAGCAAACC AAGGTAAAAA GGAATAACTG TATCCTTATA CCCCCGCAAA 7500

ATTCCCTGAA GCGGCGCCGC AAAGGTATCT GCTAACTGGA AGAAAAGACT ATAAGTTAAA 7560

AAACGCACTG TCAAATCGAT AAATTTTGGG TCGTTACCAT AAAGACTGGC CACATTTCCC 7620

CTAAAAATGT AAAGGAAGGT TAAGGTGAAG GCCGCAAAAA TGAGGGCAGT CCATCTTCCT 7680

AGACCAATAT AGGTTTTCGC ATCATCAAAT CGCTTGGCTC CCACTTCATA GGAAACGACA 7740

ATAGCCATAG CCGATGAGAT ACTCATAGGA AAGGCGTACA TAAGACTTGA AAAGTTCATA 7800

GCTGACTGGT GACTAGCTAT AATCAAGGGC GAAAACTTAG CCATAATCAA GCCAACCACT 7860

GAAAAGATAG CCACTTCCGC GAAGACAGTT CCCCCAATAG GCAGACCTAA ACGAACTCCT 7920

TCCTTAATTT TATCCATATT AAGTGGAATT CGTTTCTCAA GGTGTAAGGC TTTGAGCTTC 7980

TCCTGTTTAA ATAAAACCAG AACAGAAATC CCAAGCAAGA CCCAGTAGGC CAAGGATGTT 8040

CCTAAACCAG CACCAGCCCC TCCCAGTTCT GGAACACCAA AGGCACCGTA AATCAAGAGA 8100

TAGTTAAATC CGCTATTGAG AGGGAGTAAC AAAAGCATGA GGTACATGGA CAGTTTGGTC 8160

AAGCCCAGCG AATCCAGCAA GGAACGAATG ACGCTAAAGA GCAACAAGGG GATAATCCCG 8220

ATAGATAAAA ACCAAAGATA GCGAACCGCT ACTGCCGCTA CTGCTGCTTC TAACCCAATA 8280

TGATTCAAGA TTATTGGTGC CAAGAAAAGT ACCATCCCCA GCAAGACCAC AGATAGGCCC 8340

AAGGCCAAAT AAATAAATTG GTAAAAATCA GACGCAACTT CTTCCTTTTT GCCTCGACCA 8400

AGATGGTGAC CAATGATAGG CACCAAGGCT GACACAATCC CTGTTAGAAA TGTAAAGAAA 8460

GGATTCCAGA TACTGGTTGC CATAGATACA CCAGCCAAGT CCATAGTGTT GTATTGACCT 8520

GTCATTGCAG TATCAACAAA AGAGGCAGAA TAATTGGCAA ATTGGTAGAT CAGGATTGGG 8580

AAGAAAATTT TTAAAAATAA TACTAACTTC TCTCGTAAAC ACTTTGTCTT ATACATACTT 8640

CTCTTTCTAT TCTGATTTAT CTAAACCAAA GAGTTTCAGA CCATAGTTTT TCAAACTTAG 8700

CGGAGGTTTA TTAGATTTTG AAGTAGTATG CCAACACGCA CATGTACGAC AATAATAGCT 8760

TCTAACTAAA CCTCCGTTAT CATATTGAAC CGCATGGTCA GCTTTTTCTT TAGTTTCATA 8820

TTGAATTTTG GAACGATTAG CTGCGGGACA GTAAATTCCA CTATTAGATT TCGCTTGTCT 8880

CTCCCTACGT TTTCGAAAAT AATTCATATT CTAACTCCTA TCAAGCTTGA TAGACGATTT 8940

GTCCCTTACA GATGGTATAT TTAACCTGCC CTTTTAAGGT TTCACCGATG AATGGTGAAT 9000

TAGCTGCTTT GGAAGCAAAA TGGGAGTCCA CAAAGCGGTC AGCCTTGGCA TCAAAAATAG 9060

TGATATCTGC TGGACCATTC TCAGCCAAGT AACCTGCTTC AAAGTTGTAA AGCTTGGCTG 9120

GGTTGTATGT CATTTTTTCA AGTAATTCCA TCAAGCTCAA CTCACCAGCT TCTACTAAAT 9180

AGGTCAAGCT GAGAGACAGG GATGTTTCTA AGCCAGTCAT ACCAGATGGC GCTTTGGTAA 9240

TATCCTCAAC ATTTTTTTCA TCTACATGAT GAGGCGCGTG GTCAGTCGCA ATAACTGTGA 9300

TGACACCTGA TTTGAGACCT TCGATAACGG CACGACGGTC TGATTCCAAA CGAAGCGGTG 9360

GATTCATCTT AGCATTGCTA CCTTGTGTTA AAAGAAGTGC TTCTGTCTTA GAGAAATGCT 9420

GTGGCGCTAC TTCTGCTGTG ACTTCTGCAC CTAACCCCTG AGCAAACTCC ACTACTTTAA 9480

CACTTTCTTC CTTAGACAAA TGCTGGATGT GAACATGGGC TTTAGTTGCA TAGGCAATCA 9540

TGACATCACG CGCCATCATA GCGTACTCAG CCACCCCAGT AGCACCGCAG ATATGGAAAT 9600

GTTCTCTAGC AATATTTTCA TTAAAGCCAA GAACACCGTT CAAACCTGGA TCTTCCTCAT 9660

GAAGGCTGAT AAAGGTATTG AGTTTTTTGG CTTCCTCCAT GGCTTCCTTG ACAATCTTAC 9720

TGCTCTCAAG CGGAATACCG TCATCAGAGA AACCAACCGC ACCAGCTTCT AAGAGTGCCT 9780

TAAAGTCAGT CAAGTTTTTA CCATTAAAGT TTTTAGTAAT GGTCGCAACT GTCTTGACAT 9840

TAATCTTCTC TTTGGCAGCT GACTGGAGAA CTGCTTGCAA AGTCTCCACG TCTGAAATGG 9900

TTGGACTGGT ATTAGCCATC ATGACGACAG TAGTAAAACC ACCTGCAGCG GCTGCTAGGG 9960

CACCAGTATG AATGTCTTCT TTATGTGTTT GACCAGGTTC ACGGAAATGA ACATGAATAT 10020

CGACCAAGCC AGGAGCAACC ACAAGACCAG TAGCATCAAT CGTTTCTGCT CCTTCTTCCG 10080

TGATCTCAGA CGCAATTTTG ATAATTTTCC CATCTTGAAC TAAGACATCA CAAACTTGAT 10140

CCAAACCAGA CTTGGGATCC ATTACACGAC CATTTTTGAT TAGTAGCATC TGCTTTCTCC 10200

CTCTTTTCTT GGACTTATAC TTCTGAGAAC ATTTCCCTTA TGCCAGCGAT TATTTCCACC 2100

GTTATTCTGG TCTTTGGTGC TCTTCTTTTA GCCTTGCCCA TAGGGATTTT TGCTGGTTTT 2160

TATCTTGTGG AATATACAAA AAAAGATTCC CTTTGTGTTA AAATCATGCG ATTGGCCTCA 2220

GATACCTTAT CTGGGATTCC TTCCATTGTT TTTGGTCTGT TTGGCATGCT CTTCTTTGTA 2280

GTCTTCTTAG GTTTTCAATA CTCTCTGTTA TCAGGAATCT TAACCTCAGT TATCATGGTG 2340

TTGCCAGTCA TTATTCGCTC AACAGAAGAA GCCCTTTTAT CTGTTAGTGA TAGCATGCGT 2400

CAAGCAAGTT ATGGACTTGG GGCAGGTAAG TTACGGACTG TTTTTAGAAT TGTTCTACCA 2460

GTTGCCATGC CAGGTATTTT AGCTGGAGTG ATACTAGCTA TTGGCCGTAT CGTTGGTGAA 2520

ACAGCTGCCC TCATGTATAC ATTAGGTACC TCTACCAATA CGCCAAGTAG TCTCATGTCT 2580

TCAGGCCGTT CTCTAGCCCT ACATATGTAT ATGCTGTCAA GTGAGGGGCT ACATGTCAAT 2640

GAAGCCTATG CTACCGGCGT GATTTTGATT ATTACTGTTT TAATGATAAA TACTCTATCA 2700

AGCTTATTAT CTCGAAAACT TGTGAAAGGA GCTTCCTAGT ATGGGAACAT TTTCAGTCAG 2760

ACACCTAGAC TTATTTTACG GGGATTTTCA AGCCTTAAAA AATATTTCGA TTCAATTACC 2820

AGAAAGACAG ATTACTGCCT TGATAGGCCC ATCTGGTTGT GGCAAATCAA CTTTTCTAAA 2880

AACCCTTAAC CGGATGAACG ATTTGGTTCC TTCTTGCCAT ATTGAAGGCC AAGTCCTCTT 2940

AGATGAGCAA GATATTTATA GTAGCAAATT CAACCTTAAT CAGCTACGTA AGCGTGTAGG 3000

GATGGTTTTT CAACAGCCTA ATCCCTTTGC CATGTCTATC TATGATAACG TGGCTTATGG 3060

CCCAAGGACA CATGGTATTC GAGACAAAAA ACAATTAGAT GCCTTAGTGG AGAAATCTTT 3120

AAAAGGGGCA GCCATTTGGG AAGAAGTCAA AGATGATCTT AAAAAGAGTG CCATGTCCTT 3180

ATCTGGCGGT CAGCAGCAAC GCCTTTGCAT TGCGCGAGCT TTAGCAGTAG AACCTGATAT 3240

TCTGTTAATG GATGAGCCGA CTTCAGCCTT AGACCCTATC TCCACTTTAA AAATTGAAGA 3300

CCTCATTCAG CAACTAAAAA AGGATTATAC GATTATCATT GTTACCCATA ACATGCAACA 3360

AGCTTCACGT ATTTCAGATA AAACTGCTTT TTTCTTAACA GGAGAAATTT GCGAATTTGG 3420

AGATACCGTT GACGTGTTTA CCAATCCAAA AGATCAGCGC ACAGAAGACT ATATTTCAGG 3480

ACGGTTCGGA TAAGGAAGGA AAAACCTATG AGAAATCAAT TTGACTTAGA ATTGCATGAA 3540

TTAGAACAAT CCTTTTTAGG ACTAGGGCAA CTTGTCCTTG AAACAGCTTC AAAAGCCTTA 3600

CTGGCCTTAG CCTCCAAAGA CAAGGAGATG GCAGAGCTAA TTATCAATAA GGATCATGCT 3660

ATCAACCAAG GTCAAAGCGC TATCGAATTG ACCTGTGCCC GTTTGTTGGC CTTGCAGCAG 3720

CCACAAGTGT CTGACCTTCG ATTTGTGATT AGCATCATGT CTTCTTGTTC AGACCTTGAA 3780

CGTATGGGAG ACCATATGGC AGGCATTGCC AAAGCTGTTT TGCAACTAAA AGAAAATCAA 3840

CTAGCCCCTG ACGAAGAACA GTTACACCAA ATGGGTAAAT TATCCCTCAG CATGCTAGCC 3900

GATTTATTGG TTGCCTTTCC TTTGCACCAA GCCTCAAAAG CTATTAGTAT TGCTCAAAAA 3960

GATGAACAGA TTGACCAATA TTATTATGCC TTATCAAAGG AAATCATTGG ACTTATGAAA 4020

GACCAAGAAA CCTCAATTCC CAATGGAACT CAATACCTTT ATATCATAGG GCATCTGGAA 4080

CGCTCGCTGA TTACATTGCT AACATTTGTG AACGCCTAGT CTACCTAGAA ACAGGAGAAC 4140

TAGTGGATTT GAATTAATTC AACTAATCCT TAAAAGAGAA GAGTACGATT AAGTACTCTT 4200

TTTTATGGTT GTAAAAAAGT TCATTTGACC AATTTAAGCA GTGTAGATAG TGAGGAGTTG 4260

TTTCAATTCT ATCGTGAACG AGGGAATGCT GAAAACTTTA TCAAAGAAAG GAAAGCAGGA 4320

TTCTTTGGGG ATAAGACAGA TAGTTCGACC ATGATTAAGA ATGAAGTACG TATGATGATG 4380

GGCTGTCTGG CTTATAATCT CTACCTCTTT TTAAAGCAGC TAGCTGGTGA TGAAGTAAAG 4440

TCCTTGACTA TCAAGCGTTT TCGACGTCTC TTCCTTCATA TTGCCGGAAA ATATGTCTCT 4500

ACTGCTAGAC GACATATTCT CAAATTCTCA AGTCTATACG CCTATTCAAA ACAGTTTCAA 4560

GCCTTATTTG ATACAATCTG CCAGATAAAT CTGATACTCC CTGTTCCATA TAGAGCTAGA 4620

GGGCAGGGGA AAACATGCCT AACAGAATAA GTCACCTTAT TTTAAAAATC GAGCATCAAA 4680

CCAAGGGAGG AGTCTGCCCT TTTTTAGGAA AAAATCAAGA CAAATCTCCT CAATTATGTC 4740

TCGAACATCA GAAATTAAGC AAAATCACCA GAAGGACAGT ATTTCAACTA GCTTTTCTGG 4800

TAATTTTTGA ACTGTGTAGT TCGTTAGTGC CAGATATGAA TAATTTGGGA TGATAAATCT 4860

TTCTTCCTCA GGTAGCCTAT CATAATACTC TTCAAAAATC TTATCAAAAA CACTCTCTTT 4920

CTTTTGGGCG ATAGTTTCAT CTTCGTATGT AGGAGTCCTC ATCAAGAAAT ACTTCAATTC 4980

TAGGTATTCC TTATCCAACT CTATATAACT TGGCATCAAC TTGTAATCTT CAACCCCCAA 5040

ACGTTCAGCA ATATATTTTA ACTTTGTTAG TATTGGTCTG GATTCTCCAT TTTCAATTCT 5100

AATTAATTGA CGGATACTTA ATTCAGACTC ATCACCACAA AATTCTGAAC GACTGATTTT 5160

TTTAGCCAAA CGTAATCTTT TAATTTTTTC GCCAAACTCT CGCAACCTAC AAGAACTTCC 5220

TGAGTTGTTT ACCTCTATTA TAAGCATATA CTGAATCAAA CTATCTATCA GATTTCTTCT 5280

CACTTTAACT AAAGACTAAG AGTTTATCCC TTCGTCTCGG TTTTTGTGTA TTTTTCCACC 5340

ATACCCCAGT AATGCAAGTG CAAAATCCCC TAGAATATGA TAGAATAAGA GAAAGAACTC 5400

TATCAAGGAG GAAATCATGG AAAAACAAAC CGTCGCCGTC TTGGGGCCTG GTTCTTGGGG 5460

AACCGCCCTT TCACAAGTCT TAAATGACAA TGGACACGAG GTACGTATTT GGGGAAATCT 5520

TCCCGAGCAA ATCAATGAAA TTAATACACA CCATACTAAT AAGCACTACT TTAAAGATGT 5580

CGTTCTAGAC GAAAATATCA TTGCCTACAC CGACTTAGCA GAAACATTGA AAGATGTGGA 5640

TGCGATTTTG TTTGTTGTCC CAACAAAAGT GACACGACTT GTTGCCCAGC AAGTTGCACA 5700

AACCTTGGAC CATAAGGTTA TCATCATGCA CGCATCAAAG GGATTAGAAC CTGATAGCCA 5760

TAAACGATTA TCAACCATTC TTGAAGAAGA AATTCCTGAA CATCTCCGTA GTGATATCGT 5820

CGTTGTTTCA GGGCCTAGTC ATGCAGAAGA GACCATTGTG CGTGACCTAA CTTTAATAAC 5880

TGCTGCTTCT AAAGATTTAC AAACAGCTCA ATACGTTCAG AAGCTATTTA GTAATCACTA 5940

CTTCCGACTT TATACCAATA CGGATGTTAT CGGGGTTGAA ACTGCTGGTG CTCTTAAAAA 6000

TATTATTGCT GTCGGTGCTG GAGCTTTACA TGGTCTTGGA TTTGGTGATA ATGCTAAGGC 6060

AGCCATCATC GCTCGAGGTT TAGCAGAAAT CACCCGCCTA GGGGTAGCAC TCGGGGCCAG 6120

TCCATTGACC TATAGCGGCT TATCTGGTGT GGGAGATTTG ATCGTAACGG GAACTTCCAT 6180

CCACTCTCGT AACTGGAGAG CTGGAGATGC TCTCGGACGA GGAGAATCCC TAGCTGATAT 6240

AGAAGCTAAT ATGGGCATGG TAATCGAAGG AATTTCAACG ACTCGAGCAG CCTATGAACT 6300

AGCCCAAGAA CTTGGAGTCT ATATGCCCAT TACACAGGCT ATTTACCAAG TTATTTATCA 6360

CGGAACCAAT ATCAAAGATG CCATTTATGA CATCATGAAC AATGAATTTA AAGCAGAAAA 6420

TGAGTGGTCT TAACCCTCTA TAGAAAGGAT TTTTATGACA TCAAAAGTTA GAAAGGCAGT 6480

CATCCCTGCT GCTGGACTAG GAACTCGATT TTTACCAGCA ACCAAGGCCC TTGCCAAAGA 6540

AATGTTGCCA ATCGTAGACA AACCAACTAT CCAGTTTATC GTGGAAGAAG CTCTCAAATC 6600

AGGTATTGAA GATATTCTAG TTGTCACTGG TAAATCAAAA CGTTCTATTG AGGACCACTT 6660

TGATTCAAAC TTCGAATTGG AATATAACCT CAAAGAAAAA GGGAAAACAG ATCTTTTGAA 6720

GCTAGTTGAT AAAACAACTG ACATGCGTCT GCATTTTATC CGCCAAACTC ATCCACGCGG 6780

TCTCGGAGAT GCTGTTTTGC AAGCCAAGGC TTTCGTCGGA AATGAACCTT TTGTCGTTAT 6840

GCTTGGTGAT GACTTGATGG ATATCACAGA CGAAAAGGCT GTTCCACTTA CCAAACAACT 6900

CATGGATGAC TACGAGCGTA CCCACGCGTC TACTATCGCT GTCATGCCAG TCCCTCATGA 6960

CGAAGTATCT GCTTACGGGG TTATTGCTCC GCAAGGCGAA GGAAAAGATG GTCTTTACAG 7020

TGTTGAAACC TTTGTTGAAA AACCAGCTCC AGAGGACGCT CCTAGCGACC TTGCTATTAT 7080

CGGACGCTAC CTCCTCACGC CTGAAATTTT TGAGATTCTC GAAAAGCAAG CTCCAGGTGC 7140

AGGAAATGAA ATTCAGCTGA CAGATGCAAT CGACACCCTC AATAAAACAC AACGTGTATT 7200

TGCTCGTGAG TTCAAAGGGG CTCGTTACGA TGTCGGAGAC AAGTTTGGCT TCATGAAAAC 7260

ATCCATCGAC TACGCCCTCA AACACCCACA AGTCAAAGAT GATTTGAAGA ATTACCTCAT 7320

CCAACTTGGA AAAGAATTGA CTGAGAAGGA ATAACAAAAT CATTTATATA AAGATTAGCC 7380

ACACATAAAT TAAGTAAATT CTCTACTTGA ATCTACCTAT TTAATAAAAA CTAATGAAAA 7440

CGCTATACTT GTATTTGTTT TTTCATTAAA ATAAGAGTAG AATAAATTAG TATAGTAAAA 7500

CAAAAAAGCA CCGAATCGGT GCGCACTTTT TCAAGTTGTG TACGGACAAA GCCTTATTTT 7560

AACTTTGCTA TGTTGTTTCT AATGGTTCCA AAATAATAAA TAATTTTAAA TTTGACTTAA 7620

CTGTTGGAGT AGTCATGGTT AAATTAAATC AACCGAGCCG AACATAAGTT GTTTAATTTT 7680

GTGGAAGCTA TTAATAAAAA TATAATAAGG GAGAAAGATA GGTGTAATTT TAATTTTAAA 7740

GTAATTGCGG ACACTATCAA AGAAAAAGAT TATGGAGAAC AAATTTGTAG AATTTATCGA 7800

AAACAATAAA AAAGTAATCA TTTCATCAGT TGCAGTTGGT GTTGTATTGG TATTAGGGTT 7860

TGGATGGTAT TCATATAACC AACAACAAGC AGAACAACAA GCAAAAATTG TACAATTAGA 7920

AAAAGATAGC AAATCAGACA AAGAACAAGT TGATAAACTA TTTGAATCAT TTGATGCATC 7980

TTCAGATGAA TCTATTTCTA AATTAAAAGA ACTATCTGAA ACTTCACTTA AAACCGATGC 8040

AGGTAAAGAC TATCTTAATA ACAAAGTCAA AGAATCATCT AAAGCAATTG TAGATTTTCA 8100

TTTGCAAAAA GGTTTGGCTT ATGATGTTAA AGATTCAGAT GACAAATTTA AAGATAAAGC 8160

AACTCTTGAA ACAAATGTAA AAGAAATTAC AAAACAAATT GATTTTATCA AAAAAGTTGA 8220

TGAAACTTTT AAACAAGAGA ATTTGGAAGA AACTCTTAAA TCTCTAAATG ATCTTGTTGA 8280

TAAATATCAA AAACAAATCG AACTTTTGAA GAAAGAAGAA GAAAAAGCTG CTGAAAAAGC 8340

TGCTGAAAAA GCAAAGGAAT CTTCTAGTCA AAGTAATTCT TCTGGTAGTG CTTCTAATGA 8400

GTCTTATAAT GGATCTTCCA ATTCAAATGT AGATTATAGT TCATCTGAAC AAACTAATGG 8460

ATATTCAAAT AATTATGGCG GTCAAGATTA TTCTGGTTCA GGAGATAGTT CAACAAATGG 8520

TGGATCATCA GAACAATATT CATCTAGCAA TTCAAACAGC GGAGCAAATA ATGTCTACAG 8580

ATATAAAGGC ACTGGTGCTG ACGGCTATCA AAGATACTAC TACAAAGATC ATAATAATGG 8640

AGATGTGTAT GATGACGATG GAAATTACCT TGGGAACTTT GGTGGCGGCA TTGCAGAACC 8700

TAGTCAACGC TAATAACTAT TTTAGAGCTG TGTTGTTTCG AATGGTTCCA AAACACATTA 8760

AAAGCTACTC ATTTTTTAAG TAGCTTTTTT CTTATTCAAG TTTACATATT ATACTCAATG 8820

AAAATCAAAT TCAAACCACG TCAGCATCGC CTTACCGTAG GTATGGTTAC TGACTTCGTC 8880

AGTTTCATCT ACAACCTCAA AACCATGTTT TGAGCTGACT TCGTCAGTTC TATCTACAAC 8940

CTCAAAGCAG TGCTTTGAGC AACCTGCGGC TAGCTTCCTA GTTTGCTCTT TGATTTTCAT 9000

TGAGTATTAG TCGTCACAAT CCCATTCCCT TGTAGAAAAG CAAAATGGCG AGTCCTACGA 9060

ACAAGACTAC CGCTCCTAAT CTCTGGCTGG TGTTATACAT CCGTTTTTCT CCTCTAACTG 9120

GAAAGATAAC TGCTAGAAAT GCGCCACCAA CTGCACCACC GATATGGCCT GCTAGGCTGA 9180

TTCCTGGAAT CAGAACACTT CCAATAATGT TAACCACAAA AAGTGTCAGA TAGGATTGCC 9240

CTAGCTGTTG GATATAAGGA TTGCGAGTTG CATAGCGAAG AACAATAATC GCGGCAAATA 9300

GCCCATAAAG AGAGGTAGAG GCGCCTGCTG CTAAGGATTT AGGACTAAAT ACAAAAACAA 9360

AGAGATTGCC CATCATTCCT GATAAAAGAT AGAGAAAGAA AAACTGCTTA GAACCGAAAA 9420

TCTCCTCTAC CTGCCTTCCA AGATAATAAA GTGAAAGCAT ATTAACAATG AAATGTTCCC 9480

ACCCAATATG AACAAAAATG GCAGACAAGA GACGCCAAAC CTGCTCGGGA AAGAGGCGAA 9540

TAGCTGGCCC ATACATGGCT CCAAATCGAA ATAATGTATC TGCCCTGTCA AAGTTTCCGC 9600

CTGCAGTGAC CAACATTAGT AAAAATACCA AGGCCGTCAC TAAGAGGAAG AAACTCGTCA 9660

CAGGGTAACG TCTATCAAAG ATTTCCTTCA TCAATTAATA CCTCCTGAAC AGGAATATCA 9720

TGGTTTTCAG GTATAAAGTC CTGAATTTGA CAAGGATATA TCGTACTCAA AGTACGACCA 9780

GAAAAATGTT CCAGATAGCG GTCATAATAG CCTCCACCGT ATCCTATCCG ATATCCTTTC 9840

GTCGTAAAAG CCAGACCAGG AACATGAATC AAATCAATCT GAGATGCATC CACCACTTCC 9900

AAATCTCCCT GTAGCTCCAG TAAGGCAAAG AAAGTTTTTA CCAACTGTTG CGGATCATAG 9960

ACCACAAAGT CCATGCGCCC CTTGGGATAA GTTTTGGGTA TTAAAACCTT CTTGCCGTCC 10020

TTCAGCGCCT GCTCAATCAG TTCCTGCGTT TGAAACTCAT GAGAAAAAGA GAGGTAGGTT 10080

GCGATGACCT TGGCTTCTTG ATAAAAGGGG TGTTGTAAAA GCCGCTCGGT TAAAGCTTGG 10140

TCTATAGCCT GTTTTTGCTC TTGAGATATA GCCTTCATTT CATGCAAGAC TTGCTTGCGT 10200

AATTCCGATT TCATAGACAA GCCCTCTATT CTGCTGCCTT CTTTTTCAGG AAACTAGACA 10260

CCGCAGCCAC CCCAATAGCT AAGACTTCTT CCTTAGGACT CATTTGAGGG TGATGAAGAG 10320

CGTAGGGACT ATCGATACCT AGCCAAAACA TCACGCCATC AACCTTTGAA AGGAGATAAC 10380

CAAAGTCCTC GCCTGTCATA GCAGGTTCGA TATCAATCAA CTCGATTCCG TCTTTTTCGT 10440

CAAAGAAGTC CATCAGTTCA CGCGCCAAGG CTGGATTGTT CTCAACAGGT AGGTATCCAC 10500

CTTGTTTGAG TTCCACTTCG ACTTCCATAT CAAAGGCAGC TGCAACCCCT TCTGCAACTG 10560

TTTTTACCCT CTTTTGCACC AAGAGACTCA TGTCCTGTGT CAAGGCACGA ATAGTTCCAT 10620

GTAAAAAAGC TGTGTCTGTG ATGACATTGT TGGTGGTTCC AGCTTGAAAA ACGCCGAAGG 10680

TCACCACTGC TCCCTCGATT GGGTTGACAT TGCGGCTAAC AACTGACTGC ACTTGGGTCA 10740

CAAAGTAACT AGCCGCCACC AAGGCGTCAT TGGCTTCATG AGGAAAAGCT GCGTGGCCAC 10800

CTTTGCCTTT GAAACGGATC TTCACCTCGC AAGTTCCTGC AAAGAGTGTA TGAGTATTAG 10860

TCGCAATCTG GCCGACTTTC AAATCTGGAC GAACATGGAG ACCATAGAAT TGATCTGGCA 10920

ACCAATCTCC AAAAGCACCG TCCTCATACA TGAGCATACC ACCAGCTTCA TTTTCTTCAG 10980

CAGGCTGAAA TAGAAAGAGC AGATTATTCT TGGGTTGCTC CTCAAGGGCG CGCTCAAGAC 11040

AGCCTAAGGC AATGGTCATA TGAAAATCAT GGACACAGGC ATGCATGCGA CCTTGGTGTT 11100

GAGAAGCAAA AGGTAGACCT GTTTGTTCGA CGATAGGCAG GCCATCAATA TCTGTCCGCC 11160

AACCAATGGT TCGCTCCGGC TGACTTCCCT GCAGGTAGAC CAAAATCCCT GTCCGCCAAG 11220

TACGAATTTG AACAAAATCC TTGCCCGTAG TCAATTTCTC AATCACATCC AGCAAATAAG 11280

CCTGAGTCTT GAACTCCTCC AAGCCAATCT CTGGAATCTG GTGTAAATCT CGTCTAGTCT 11340

GAATCAAATC TAACATCTAT CTGTCCTCCG ATATAGCAGA AAGAGGCTGG AAAAAGGGTT 11400

CCGCCTCTTT TTTACTTTTA CAATTACAAG GTACGAAGCG CATCCTCTAG CGCTGTTTTT 11460

TGTTGAGTTT GGGCATCAAT TTCTTTGATA ATACGAGCTG GAACACCTGC TACTACCACG 11520

TTTTCTGGGA CATCTTGGGT AACAATAGCT CCTGCTGCGA CAACTGAACC ACTACCGATT 11580

TGGACTCCTT CGATAACCAC TGCATTAGCA CCGATAAGAA CATTGTCTCC GACACGGACT 11640

GGTTCAGCAC TAGCTGGCTC AATCACACCT GCCAAAACTG CACCTGCACC AACGTGGCTA 11700

TTTTTTCCAA CGATGGCACG GCCACCAAGG ATGGCACCCA TGTCAATCAT GGTTCCAGCA 11760

CCGATTTCAG CACCGATATT GATAACAGAT CCCATCATGA TAACAGCATT GTCACCAATT 11820

TCCACCTGGT CACGGATAAT CGCACCTGGC TCGATACGAG CGTTGATAGC ACGCTTATCT 11880

AGCAAAGGAA CTGCAGAATT ACGAGCATCT TGCTCGACAA CATAATCTTG ATTTTCTACC 11940

AAACCTTCAA GAAGCGGAGC CACATCCTTC CAGTCTCCGA ATAGGACATT TCCTAGTTTG 12000

ACAACAGAGC TAGGCACAGC AGTTGCGAGT TGCCCCTCAA AGGTTACTTT GACACTGGTT 12060

TTCTTTTCAG CATTGGCGAT AAATTGGATA ATTTCTTGAG CGTTCATTTT TGTAGCAGTC 12120

ATAGGTG 12127

(2) INFORMATION FOR SEQ ID NO: 149:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12566 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

CCATCCTTCT GTTGATGTGA CAGGAATGAT GATAAATCAA CCAGTAGCTA GTCGCGAAGA 60

GGTGACAGAG GCTTTGAGTC ACTTGGCGGT AGAGCACAAT AGTCTCATTG CTCGTCGAAT 120

CGTTGAGCCA AATGAAGCTG GAGAAACACG CTTTACCTAT GCCACTTATG GTGAGGGAAA 180

GCTTCCAGAA GGTCTGACCA TTTCCTCCAA GGAGAGTGCA GAAACGAGTG ATTTATTAGG 240

GTCTTACTTG ATTGTATCAG GAAGTTTGGA TGGAGTGAGC TTACAGACCA CCTTGAAAGA 300

GCTTGGTTAT CAAGGCTTTG TTTCGAATGG AGAAGATCCA TTTTCGATAG TCTTACTATT 360

GACGGCCACC CCTATGGTGC TACTGAGTTT AGCTATTTTT CTGCTGACCT TTATGAGTCT 420

GACCCTGATT TATCGGATCA AATCCCTTCG TCAGGCAGGG ATTCGCTTAA TAGCTGGTGA 480

GAGCTTGTTT GGAGTTGCTC TCAGACCAGT GTTAGAAGAT GTGAGACAGC TTATCTGCTC 540

AGTGCTGGTA TCCAGTCTTT TGGGATTGGG GATTCTCTGG TATCAAGGTG CCTTGTTTAT 600

GGCAACGGTG CAACTGGTCA TCATTGCTCT TCTACTTTAT GGATTGACCT TGGCAGGGAT 660

TTCTACCTTA CTAAGTGTCG TCTATCTACT TGGTTTACAG GAAAATAGTC TGGTGGATCT 720

ATTGAAAGGG AAACTCCCTC TCAAACGTAT GATGACATTG ATGATGGTGG GGCAACTCTT 780

AGCTGTATTG GTGGTCGGAT CGAGTGCGAC AGCTCTCCTA CCCCACTACC GTGAAATGCA 840

GGAAATGGAG AGAGCTAGCA ATAAATGGAG CCAGTCCTCA GACCGTTACC GTCTATCCTT 900

TGGTTGGTCT AGTGCATTTG CCGATGAAGA AGGAACGCGT AAGGATAATC GTGAGTGGCA 960

GACATTTACT GAAGAACGGT TAGCCAATAC AGACTCTTTT TATATTATGA GCAATGTTGA 1020

CAATTTCTCA GATGGAGCAG AAGTGGACCT AGATGGCAAT CGTCTCAGTG ACTACACACC 1080

GTCAGGGAAT GTTATCTATG TCTCACCGCG CTATCTGATA GAAGAAAAGA TTACCGTTTC 1140

TTCAGAGTTT ATGGACAAGA TGCAAAACTT GTCTGAGGGA GAGTTTGGGC TGATCTTGCC 1200

TGAGAGCTTG CGAGAGCAGT CTGTCTACTA CCAAGGATTG TTTACAGATT ACCTGCAAAA 1260

CTTTTCATCT GAAAGTGTAG AAGTGACGAG TCAGAAACAC TACCTCCCAC AGGTAAGGCT 1320

AGCTTTTACA GAAACAGGAC AGGAACGTTT CCTCTATAAT GATGGGTACA AGACAACACG 1380

CCAGTACCTA AAAGATCCGA TTATTGTAGT TCTAACGCCG CAAGCGACTG GAACAAGACC 1440

TGTTGCAGGG ATGTTGTGGG GAACTACGGC TAATAGTGCC TTGAAACTAG ATCGATATGG 1500

AGACAGCATC ACAGCTCTAA AAGAGAAAGG TCTGTATCAC AAGGTTTCTT ACTTGGTAAA 1560

AAGCCAGCTA TTTTTTGCCA AGGTACTAAA TGACAAACGG GTGGAGTTTT ACTCTCTCCT 1620

TATTGGGACG ATTTTGACCC TGTCTACGGC TATCTTGTTA TTTGATTCCA TGAATCTTCT 1680

CTATTTTGAG CAGTTCAGAC GGGAACTTAT GATTAAACGT CTTGCTGGTA TGACAATCTA 1740

TGAGCTTCAT GGCAAGTATT TACTGGCGCA AGGAGGAGTT CTCTTGCTTG GCCTAGTCCT 1800

ATCTAGTATT TTGACAAGAG ATGGTTTGAT TAGCGCTCTA GTTGTAGCTT TGTTTACGCT 1860

1000

TAACGCCCTC TTGATTTTAG TAAGGCAGGA CAAAAAAGAA GAAGCTGGTA GCATGGCAGT 1920

ATTGAAAGGA AAATAAGATG ATTGATATTC AAGGATTGGA AAAGAAATTT AATGACCGCG 1980

CGATTTTCTC TGGTTTGAAT CTCAAGCTGG AGAAGGGCAA GGTTTATGCC TTAATCGGAA 2040

AGAGTGGAAG CGGAAAGACG ACGCTGCTGA ATATCTTGGG AAAGCTAGAA AAGATAGATG 2100

GTGGAAGGGT TCTCTATCAG GGGAAAGATT TAAAAACCAT TCCCACTCGT GAGTATTTTC 2160

GAGACCAGAT GGGCTATCTC TTTCAAAATT TCGGCCTCTT AGAAAACCAA TCAATCAAAG 2220

AAAATTTGGA TTTGGGTTTT GTTGGTCAGA AAATCTCAAA AGTAGAACGT TTGGAAAGGC 2280

AAGTGGGGGC TTTAGAAAAA GTTAATCTAG GGTATTTGGA TTTAGAACAA AAAATCTATA 2340

CTTTATCTGG GGGAGAGGCC CAACGAGTTG CCCTTGCTAA GACTATTTTG AAAAATCCAC 2400

CCTTGATTTT GGCAGATGAA CCAACAGCAG CTCTTGATCC TGAAAATTCA GAGGAGGTTA 2460

TGAATCTCTT GGTGGATTTG AAAGATGAAA ATCGAATTAT CATCATTGCG ACCCATAATC 2520

CCCTAGTCTG GAATAAGGCT GATGAAATCA TTGATATGAG GAAACTTGCT CATGTGTGAA 2580

AAAATCCGTA TTCGCAGGGT ATCTGATTAT CCTAGTGCCA GAGGTGGTTT AGAAGATATC 2640

CTCATCATGG AAAATATGAC CAATCATCTC CTTTTGGTTC AAATCCGAGT GCATGGCTAT 2700

TTGCTTGATT TTGCTAGTAT TGAAGGGCAA AGGCAAAAGC ATTATCGTTT GAAAAATTTA 2760

CCTCAGACGG TTGAACTGAC AGTGGATGAT GTGGAGGAGG ATGTGGATTT GACCCTACCT 2820

GAAAATCGAA GTTATCAAGA AGCTGATTTT TTTGAACGCA TGTTTCGAGA GAACTGCTAA 2880

GGCCACTTTT AAAGATTTCC AAGACTATCT TTCTTCATGA GGAAAGATAG TTTTTTGGTA 2940

TGATTTTCAT TCCCAAAATA CAAGGGGAAT GTGTTACAAT AGTAGTAACA GATAATAGAA 3000

AAGAGAATAG ATGAGAATTG CAGATTATAG CGTGACCAAG GCAGTGCTGG AGCGTCACGG 3060

TTTTACCTTT AAAAAGTCCT TTGGGCAAAA TTTTTTGACG GATACCAATA TCCTTCAAAA 3120

AATTGTGGAT ACGGCTGAAA TTGATGATCA GGTCAATGTC ATCGAAATCG GGCCAGGTAT 3180

TGGTGCCTTG ACAGAATTTT TGGCTGAGCG TGCAGCCCAA GTCATGGCTT TTGAGATTGA 3240

CCACCGTTTG GTGCCAATTT TGGCAGATAC CCTGCGTGAT TTTGATAATG TGACCGTAGT 3300

TAACGAAGAT ATTCTCAAGG TTGATTTGGC GCAACATATC CAGAATTTTA AAAATCCTGA 3360

CCTGCCAATC AAGGTAGTGG CTAATTTGCC TTACTACATC ACGACGCCTA TTCTCATGCA 3420

CTTGATTGAG AGTGGCATTC CTTTTTGTGA GTTTGTGGTC ATGATGCAGA AAGAAGTAGC 3480

GGACCGCATT TCAGCCCAGC CTAACACCAA GGCTTACGGT AGCTTGTCTA TCGCCGTGCA 3540

GTATTACATG ACAGCCAAGG TTGCCTTTAT CGTGCCTCGT ACGGTCTTTG TGCCAGCGCC 3600

AAATGTGGAT TCAGCCATCT TGAAAATGGT GCGTCGTCCA GAGCCAGCCG TAGCAGTAGA 3660

1001

AGATGAGAAC TTTTTCTTTA AGGTTTCCAA GGCTAGTTTT ACCCATCGCC GCAAGACCTT 3720

GTGGAATAAC TTGACAGGTT ACTTTGGTAA GACTGAAGAG GTCAAGGACA AGCTGACCAA 3780

GGCTTTGGAC CAGGCAGGCT TGTCACCAAG TGTGCGTGGG GAAGCTCTCA GCTTGGCAGA 3840

ATTTGCCGGT CTAGCAGACG CACTTAAAGG GCAAGGACTC TAAGATGCAG GGACAAATCA 3900

TTAAAGCCTT GGCAGGTTTC TACTATGTGG AGAGTGATGG CCAGGTTTAT CAAACACGCG 3960

CGCGTGGGAA TTTCCGTAAA AAAGGCCATA CCCCTTATGT TGGGGACTGG GTAGATTTCT 4020

CTGCCGAGGA AAATTCAGAA GGCTATATCC TCAAAATTCA CGAACGGAAA AACAGTCTGG 4080

TTCGTCCGCC TATTGTCAAT ATCGATCAAG CTGTAGTAAT CATGTCCGTC AAGGAACCTG 4140

ATTTTAACAG CAATTTGCTG GATCGTTTCT TGGTTCTTTT GGAGCACAAG GGCATCCATC 4200

CCATTGTCTA TATTTCCAAA ATGGATTTGT TGGAAGATAG GGGAGAACTG GATTTTTACC 4260

AGCAGACCTA TGGTGACATC GGCTATGACT TTGTGACCAG TAAAGAGGAA CTCCTGTCTT 4320

TGTTAACAGG CAAGGTTACG GTCTTTATGG GGCAGACAGG TGTTGGGAAG TCAACTCTTC 4380

TCAATAAAAT CGCACCAGAC CTCAATCTTG AAACGGGAGA AATTTCAGAC AGTCTAGGTC 4440

GCGGTCGCCA TACCACTCGA GCTGTTAGTT TTTACAATCT CAACGGGGGT AAAATCGCAG 4500

ATACACCAGG ATTTTCATCC TTGGACTATG AAGTATCAAG GGCTGAAGAC CTCAATCAGG 4560

CTTTCCCAGA GATTGCTACT GTTAGCCGAG ATTGTAAGTT CCGTACTTGT ACCCATACCC 4620

ATGAGCCGTC TTGTGCCGTC AAACCAGCTG TTGAAGAGGG TGTTATTGCA ACCTTCCGTT 4680

TTGACAATTA CCTGCAATTC CTTAGTGAAA TTGAAAATCG TAGAGAAACC TATAAAAAAG 4740

TCAGCAAAAA AATTCCAAAA TAAGGAGAAA CCTATGTCTC AATACAAGAT TGCTCCGTCA 4800

ATTCTGGCAG CAGATTATGC CAACTTTGAA CGTGAAATCA AACGTCTAGA AGCAACTGGG 4860

GCAGAATATG CCCATATCGA TATCATGGAC AGTCATTTTG TACCGCAAAT CAGTTTTGGT 4920

GCAGGTGTGG TCGAGAGCCT TCGTCCTCAT AGTAAGATGG TTTTCGATTG CCACTTGATG 4980

GTGTCAAACC CTGAGCATCA TCTGGAAGAT TTTGCGCGTG CAGGTGCAGA CATCATCAGT 5040

ATCCATGTAG AAGCAACGCC TCATATTCAT GGCGCCCTCC AAAAAATTCG TTCACTCGGA 5100

GTTAAGCCTT CAGTCGTTAT CAATCCTGGC ACATCAGTTG AAGCCATCAA GCACGTCCTT 5160

CATCTAGTTG ACCAAGTTTT AGTCATGACG GTTAATCCAG GTTTTGGTGG GCAAGCCTTT 5220

CTGCCAGAAA CCATGGATAA GGTCCGTGAG TTGGTTGCTC TTCGTGAGGA AAAAGGTTTG 5280

AACTTTGAAA TCGAAGTGGA TGGTGGGATT GATGACCAAA CTATTGCTCA AGCCAAAGAA 5340

GCCGGTGCGA CTGTTTTTGT AGCAGGTTCC TATGTCTTTA AGGGAGAAGT CAATGAGCGA 5400

GTACAAACTC TCAGAAAACA ACTGGACTAG GGTTGCAGTT TTTGCAGGCG GAAACCGCGG 5460

TCATTATCGG ACAGATTTTG ATGCTTTTGT TGGGGTGGAT CGAGGCTCGC TCTGGGTCTT 5520

GGAAGAAGAC TTACCTCTTG CTCTAGCAGT CGGAGATTTT GATTCTGTGA CGGAAGAAGA 5580

GCGACAGGTG ATTCAAAAAG GTGCCCAGTA TTTTGTCCAA GCACGACCAG AAAAGGATGA 5640

TACAGATCTG GAATTGGCTC TCTTAACCAT CTTTGAACAA AATCCTCAGG CTCAGGTCAC 5700

TATTTTCGGT GCCTTGGGTG GCCGTATTGA CCATATGTTG GCCAATGTCT TTCTGCCTAG 5760

CAATCCTAAG TTGGCACCCT ATATGCATCA AATAGAAATT GAGGATGGGC AAAACTTGAT 5820

TACTTATTGT CCAGAAGGAA TCAGTCAGCT AGAACCTCGT TCAGACTACG ACTATCTAGC 5880

CTTTATGCCA GTTCGGGATA GCCAGCTGAC TATTCTTGGA GCCAAGTATG AGTTGACAGA 5940

GGAAAATTTT TTCTTTAAAA AAGTGTACGC TTCTAACGAA TATATAGATA GGGAAGTGTC 6000

GGTAACTTGC CCAGATGGTT ATGTGGTCGT ACTGCATAGC AAGGACAGGA GGTAGGATGG 6060

AAAGTTTACT TATTCTATTA TTAATTGCCA ATCTAGCTGG TCTCTTTCTG ATTTGGCAAA 6120

GGCAGGATAG GCAGGAGAAA CACTTAAGTA AGAGCTTGGA GGATCAGGCA GATCATTTGT 6180

CAGACCAGTT GGATTACCGC TTTGACCAAG CCAGACAAGC CAGCCAGTTA GACCAAAAAG 6240

ATTTGGAAGT GGTTGTCAGC GACCGTTTGC AAGAAGTGCG GATTGAATTG CACCAAGGTC 6300

TGACCCAAGT CCGTCAAGAA ATGACAGATA ATCTCCTCCA AACTAGAGAC AAGACAGACC 6360

AACGTCTCCA AGCCTTGCAG GAATCAAATG AGCAACGTTT GGAACAAATG CGCCAGACGG 6420

TCGAGGAAAA ACTAGAAAAG ACCTTGCAGA CACGCTTACA GGCTTCCTTT GAGACAGTTT 6480

CTAAACAACT GGAGTCTGTC AATCGTGGCC TTGGAGAAAT GCAGACAGTT GCCCGTGATG 6540

TCGGAGCTCT TAACAAGGTT CTCTCTGGAA CCAAGACGCG AGGGATTCTG GGAGAATTGC 6600

AACTGGGGCA AATTATTGAA GACATCATGA CACCTGCCCA GTACGAACGA GAATACGCAA 6660

CGGTTGAAAA CTCTAGTGAA CGAGTGGAGT ATGCCATCAA GTTACCCGGA CAAGGCGACC 6720

AAGAATACGT CTATCTGCCA ATTGACTCTA AGTTTCCACT GGCAGATTAT TACCGCTTGG 6780

AAGAAGCCTA TGAGACAGGT GACAAGGATG AGATTGAACG CTGTCGTAAG TCACTCCTAG 6840

CAAGCGTCAA GCGCTTTGCT AGGGATATTA GGAACAAGTA CATAGCACCA CCTCGGACGA 6900

CCAATTTTGG AGTTTTGTTT GTTCCGACAG AAGGTCTCTA CTCAGAAATC GTCCGCAATC 6960

CGGTCTTCTT TGATGATTTG AGACGGGAAG AACAGATTAT TGTTGCAGGA CCAAGTACCC 7020

TATCAGCCCT TCTTAACTCC CTATCAGTTG GTTTCAAGAC CCTTAATATC CAAAAGAGTG 7080

CCGACCATAT CAGCAAGACT CTTGCCAGTG TCAAGACCGA GTTTGGCAAG TTTGGTGGTA 7140

TTCTGGTCAA GGCACAAAAA CATCTCCAAC ATGCCTCTGG CAATATTGAT GAATTATTAA 7200

ACCGTCGTAC CATAGCTATC GAGCGGACGC TCCGTCACAT TGAGTTGTCA GAAGGTGAGC 7260

CTGCGCTTGA TCTACTCCAT TTTCAAGAAA ATGAGGAAGA ATATGAAGAT TAGTCACATG 7320

AAAAAAGATG AGTTATTTGA AGGCTTTTAC CTAATCAAAT CAGCTGACCT GAGGCAAACT 7380

CGAGCTGGGA AAAACTACCT AGCCTTTACC TTCCAAGATG ATAGTGGCGA GATTGATGGG 7440

AAGCTCTGGG ATGCCCAACC TCATAACATT GAGGCCTTTA CCGCAGGTAA GGTTGTCCAC 7500

ATGAAAGGAC GCCGAGAAGT TTATAACAAT ACCCCTCAAG TCAATCAAAT TACTCTCCGC 7560

CTGCCTCAAG CTGGTGAACC CAATGACCCA GCTGATTTCA AGGTCAAGTC ACCAGTTGAT 7620

GTCAAGGAAA TTCGTGACTA CATGTCGCAA ATGATTTTCA AAATTGAAAA TCCTGTCTGG 7680

CAACGGATTG TCCGAAATCT CTACACCAAG TATGATAAGG AATTCTACTC CTATCCAGCT 7740

GCCAAGACCA ACCACCATGC CTTTGAAACG GGCTTGGCCT ATCATACGGC GACCATGGTG 7800

CGTTTGGCAG ACGCTATTAG CGAAGTTTAT CCTCAGCTCA ATAAGAGCCT GCTCTATGCG 7860

GGGATTATGT TGCATGACTT AGCTAAGGTC ATCGAGTTGA CGGGGCCAGA CCAGACAGAG 7920

TACACAGTGC GAGGTAATCT TCTTGGACAT ATCGCTCTCA TTGATAGCGA AATTACCAAG 7980

ACAGTTATGG AACTCGGCAT CGATGATACC AAGGAAGAAG TCGTTTTGCT TCGTCATGTC 8040

ATCCTCAGTC ACCACGGCTT GCTTGAGTAT GGAAGCCCAG TCCGTCCACG CATTATGGAA 8100

GCAGAGATTA TCCATATGAT TGACAATCTG GATGCAAGCA TGATGATGAT GTCAACAGCT 8160

CTTGCTTTGG TGGATAAAGG AGAGATGACC AATAAAATCT TCGCTATGGA TAATCGTTCC 8220

TTCTATAAAC CAGATTTAGA TTAATAATTT AAGAAAAATG AGCATTTTTT AGGATAAGAA 8280

TGTTCGTTTT TTTATGTGAA TATGGTATAA TAAGTAAAAG ACAAAAATGA ATACTCTTCG 8340

AAAATCTCTT CAAACTAGGG TAGTATCGCC TTGTCGTATG TATATATGCA GGTATATTAC 8400

AGGGTTTGTC AGTTCTATTG ACAATCTCAA AACAGTGTTT TGAACCACCA GCGACCAGCT 8460

TTCTAGTTTG CTTTTTGATT TTTTGAATAA AAATGGAATA GGAAATAGAA ATGAAATTAA 8520

GAAGAAGTGA TCGGATGGTT GTCATTTCCA ACTATTTGAT TAATAATCCT TATAAACTAA 8580

CTAGTCTCAA TACTTTTGCT GAAAAGTATG AGTCTGCTAA ATCATCCATC TCAGAAGATA 8640

TCGTCATTAT CAAACGCGCC TTTGAGGAAA TTGAAATCGG TCATATCCAG ACAGTGACTG 8700

GGGCTGGCGG AGGTGTCATC TTCACACCGT CTATTTCGAG TCAGGATGCT AAGGAAATGG 8760

TTGAAGACTT GCGTACCAAG TTGTCAGAAA GTGACCGTAT CTTGCCAGGT GGTTATATCT 8820

ATCTGTCTGA TTTGCTTAGC ACACCAGCCA TCTTGAAAAA TATTGGTCGT ATTATTGCCA 8880

AAAGCTTTAT GGACCAAAAA ATTGACGCGG TTATGACCGT AGCAACTAAG GGTGTGCCAC 8940

TTGCAAATGC AGTTGCCAAT GTCCTCAATG TCTCTTTTGT CATTGTGCGC CGTGACCTGA 9000

AAATTACCGA AGGTTCAACT GTTAGCGTCA ACTATGTTTC AGGTTCAAGT GGTGACCGTA 9060

TCGAGAAAAT GTTCCTTTCA AAACGTAGTC TTAAGGCAGG CAGCCGTGTC TTGATTGTGG 9120

ATGACTTCTT GAAAGGTGGC GGAACGGTCA ATGGTATGAT TAGTCTCTTG CGCGAGTTCG 9180

ACTCAGAACT GGCAGGTGTA GCGGTCTTTG CGGACAATGC CCAAGAAGAA CGTGAAAAGC 9240

AGTTTGACTA CAAGTCACTC TTGAAGGTAA CCAATATTGA TGTCAAGAAC CAAGCCATCG 9300

ATGTTGAGGT TGGCAATATC TTTGACGAAG ATAAATAAGA GATAGAACTA AAGGTTGGAA 9360

CGATTGTCCC AGCCTTTCTT TGCAAACAGA ATAGAAGGAA GCTTATGAAA ACACCATTTA 9420

TCAATAGAGA AGAGTTAGAA GCGATTGTTG CCGAGTTCCC GACTCCCTTT CACTTGTATG 9480

ATGAGAAGGG GATTCGTGAG AAGGCAAGAG CCGTCAACCA AGCTTTTTCG TGGAACAAGG 9540

GCTTTAAGGA ATATTTTGCA GTTAAGGCTA CTCCAACTCC AGCTATTTTG AAAATTCTCC 9600

AAGAAGAAGG TTGTGGTGTG GACTGCTCTA GTTATGTAGA GCTTTTGATG AGCCATAAAC 9660

TGGACTTTCT GGGTTCTGAG ATTATGTTCT CTTCCAACAA CACGCCAGAC AAGGAATACG 9720

CCTATGCACG TGAATTGGGT GCGACCATTA ACTTGGATGC CTTTGAAGAT ATTGAACATC 9780

TGGAGAGAGT AGCAGGCATT CCAGAAATCA TCTCTTGTCG TTATAATCCT GGAGGCGTTT 9840

TTGAACTGGG GACAGACATT ATGGACAATC CTGGGGAGGC TAAGTTTGGC ATGACCAAGG 9900

ACCAGCTCTT TGAAGCCTTT GCTATCTTGA AGGAAAAAGG AGCCAAGACT TTTGGGATTC 9960

ACTCCTTCCT AGCGTCCAAT ACCGTGACCC ATCTCTATTA TCCAGAGTTG GCTCGTCAGC 10020

TCTTTGAACT GGCTGTTGAA ATCAAGGAAA AGTTGGGCAT TTCGCTAGAC TTTATCAATC 10080

TTTCTGGCGG TATTGGTGTT AATTATCATC CAGACCAGGA GCCGAACGAT ATCGCCTTGA 10140

TTGGTGAGGG AGTTCGTAAG GTGTATGAAG AGGTTCTTAC GTCAGCAGGT CTTGGTCAGG 10200

TCAAGATTTT CACCGAATTG GGTCGTTTTA TGCTGGCACC TCACGGTGCT CTAGTCACAA 10260

GAGTCACTCA TAAGAAGGAA ACCTACCGTA CCTATCTAGG TGTGGATGCC TCAGCAGTCA 10320

ACCTCATGCG TCCAGCTATG TACGGAGCTT ACCATCATAT TAGCAACGTG ACCCATCCAG 10380

ATGGACCAGC TGAAGTGGTA GATGTGGTCG GTTCACTCTG TGAAAACAAT GATAAATTTG 10440

CAGTTAATCG CGAACTGCCT CATACAGAAA TCGGTGATTT GCTGGTCATT CATGATACAG 10500

GTGCCCACGG ATTTTCAATG GGCTACCAGT ATAATGCCAA ATTACGTTCT GCGGAAATCC 10560

TCTATACCGA AGAAGGTAAA GCCCGTCAAA TCCGCCGTGC AGAGCGCCCT GAGGACTATT 10620

TTGCAACCTT ATATGGCTTC GATTTTGAAG AATAATCTGA TAATAGATTG AAAATGAAAT 10680

TGAAAAACAG ATTGCTTTCT AAAAAATAGG CAAAAATCTT GTTTTTCCTT CAAGTCGTGA 10740

TATAATAAAA CTATAAAACG TTTTCAAGGA AGGTAACGAT ATGTCTGAAG AAACAATTGA 10800

TTATGGACAA GTGACAGGAA TGGTGCATTC GACAGAAAGC TTTGGGTCAG TAGATGGGCC 10860

TGGTATTCGC TTTATTGTCT TTTTGCAGGG CTGTCACATG CGTTGCCAGT ATTGCCACAA 10920

CCCAGACACT TGGGCTATGG AGTCCAATAA GTCACGTGAA CGGACGGTAG ATGATGTCTT 10980

GACAGAGGCC TTGCGCTACC GTGGTTTCTG GGGAAATAAG GGTGGGATTA CAGTCAGTGG 11040

AGGAGAAGCT CTCTTGCAGA TTGATTTCCT GATTGCTCTC TTCACCAAGG CTAAGGAACA 11100

AGGAATCCAC TGTACCTTGG ACACCTGTGC TCTTCCTTTC CGTAATAAAC CACGTTACCT 11160

TGAGAAGTTT GACAAACTCA TGGCTGTCAC TGACTTGGTT CTTTTGGATA TCAAGGAAAT 11220

CAACGAAGAA CAGCACAAGA TTGTCACTAG CCAAACCAAT AAAAATATCT TGGCTTGTGC 11280

CCAGTATCTA TCAGATATTG GAAAACCTGT CTGGATTCGC CACGTGCTAG TTCCAGGATT 11340

GACAGACAGA GATGATGACT TGATTGAACT TGGTAAGTTC GTCAAGACCC TCAAAAATGT 11400

TGATAAGTTT GAAATTCTAC CTTATCACAC CATGGGTGAG TTCAAGTGGC GTGAACTTGG 11460

AATTCCATAT TCCCTCGAAG GAGTCAAACC ACCAACAGCA GATCGCGTCA AGAACGCTAA 11520

ACAACTCATG GATACCGAAA GTTATCAAGA TTATATGAAA CGTGTACATG GATAGAAAAG 11580

AAGCCTGATG GAAACATCGG GCTTTTGACT TGCAAAAAGA CTTAGCAAAT CAGCTAAGCC 11640

TTTTTCTTCT TATCTCGAAC GTTGTTTTCC AGCGTTGCGA TTTTTGTGTT TTTTCTTGCT 11700

TGTGATAGCA GTTGGTTGTT CAGGGGTAAC GTCTTTTCGT CCACTTGGTT TAGAGAAAGC 11760

ACTTGCTTTT GGTGGGTTCT TGGCTAGTTC TTCACGGACT TTTTTGCGAA GTTTTGGACG 11820

AACGATATAG TTGACGATAA ACTGTTGGAG AATCATCATG AAACCACCGA CAACCCAGTA 11880

AAGTGTGACA CTAGCTGGTG AGAAGAGGGA GAAGACGACG ATCATGAGTG GGCTCATGTA 11940

AATCATTTTC TTGATTTGTT CTCTTTGCAT TTCATCTTCT ACTCCGTGAA GTGAAAGGAG 12000

CGATTGAAGA TAGTAAAGGA CACCAGCACA GGCAACCAAA ATCATACTTG GAGAACCTAG 12060

AGGAATGCCT AGGTAGCTTG CTTGAGCAAC CCCTTCAGTA TGTTGGGCAG CAAAGTAGAT 12120

AGCAGAGAAG AAAGGCATTT GAAGGAGGAT AGGGAAACAT CCTACACCGC CAAACATGCT 12180

GATACCGTGC TCTTTTTGAG CAGCAAAGAG AGCTTGTTGG GCTTCGAGTT TTTCTTCTTG 12240

AGTAGTCGCT TCTTTGAGAC GCGTTTGGTG TGGCTCAAGG ACGTGCTTGA GGGCGTTCAT 12300

CTTTTCAGAG TGAAGCGTTG CCTTCCATGA TTGGTAGATA CCAAGTGGTA AGATAATCAA 12360

GCGTACGATA ATGGTTACGA TAATGATAGC GACACCAAAG CCTAGACCTT TATCAGTAGC 12420

GAAGTACTTG ATGGCTTCAG CCATAGGCGC TCCGATCGTA TTCCAAATAA ATCCTGTTGG 12480

CTGACCTGTG GTTTTATCGA CATTGACACA GCCAGTCAAG ACAAGCAACA TAGCCACTCC 12540

CATAGCCGAG AGTGCAAAAT CGGGGT 12566

(2) INFORMATION FOR SEQ ID NO: 150:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5238 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

TGACACTCTG TAGGATTGTC GTTAATTGAT TGCTCGTACT CTCTACAATA ACCACCAAAG 60

TAAAAACGAC ATAGAAAGAT AGCATCAGCT GTAGCCATAG CGCCTTTGAC ACCTTCTGGA 120

TGATTATGAG TTACCTCTGC AGAAAGACTC GTAAGTCCTC TAGATGATGG CCATATACCA 180

GTTTTCGCAT AAAAACCACA GTCCATGATC CAAGCACATG GAGAAATACG CATAGCTGAT 240

CCATTCCCAA AGCTATTATA AGGCTCACGG TTATCGCTGT TTAGCCATGC ATTAAACCGA 300

GCACCGTAAT CAGCATTCGG ATACATTCTG CCATATTTCT TCATCGCGTC AATGAAGTCA 360

TCTTTTTGTC CACCATTCAT AATTGCTTCT GCAACAGCAC AGGTCATAAC CGTGTCATCT 420

GTAAAAAAGC AGTCCTTCCG AAATAAAGGA AAGTCCTTTG TTTTGATATT GTTCCATTCG 480

TAAACAGAAC CGACAATATC TCCAATAATT GCTCCAAGCA TCAGATTCCT CCTTGTTCAT 540

TTTGATGCTT TTTATATTGG TTATCTACCA TATTTATTTT AGAAAATAAC ATCCTGTTGG 600

ATTTTAAAAA TTTCATTTTT TTCAAAATAG GGTTTTACCA TTTCTTTCCA CCTAGCTCTA 660

TGAAAATTGA TTGATTTTAA AGGAGATAGG CCATAATTTC CCAATGCATA ACCATCATTT 720

ACTTCAACAA CAAGTGTTCT GCCATCGCGA GTAACACCGA TATCTAGTCC ATAAGCTATT 780

GGCGCATCTT TCCAACATGA TATCGCTTCA TCAATTACAC TTGCATCAAA TTGTGCATGA 840

TAATCACCTG TATAGGGTCG AACATCTAAT ACGCGACCAT CTAACACAAA ACAACGCCAT 900

TCAGCTATGA ATTCTACAAC CTCACTAATC CATATAGGAT AGTCGAAAGG TAGACCAATA 960

CCTATTAAAT CATGGGTTCC ATTAACAACT CTTCCAGTAA AGACTTTTGA ACCAGCTTTA 1020

GGCTTAATAA ATTTTCCCCA ATTATCAGGT ATATTCACAA TCTCTCCTAA AATACCAGCA 1080

TAAATCTTTC GACCATAAAA CTCTTTAAGC TCAATAGGAT AGTCATGAAC CGGAACGTTT 1140

AAGCCCATCA TTTTTAGTAA TGCTCTAGTC TCCATTATAT AATCTACAAC TATATCTTCA 1200

CTTGTTAACT CTTTTATTTC AGAAAAAGAT TGATATAAAA TAACTTCTTC TCCTTGTAAG 1260

TAGGCACCTA CTTGAGCATT GTATTTATTA ATTGAAACCT CACTTGGTAA TTTACTTTGT 1320

CTAATATAAA CAACCATTTC ATCACTCCTA TATCACTAGT GTTACACCAA TTTGTAAAAA 1380

ATAATAGCAA TTTTGCTCTT ATTTTTTTGA GTAAATAGCC CCCATAATAT CATCGAAATA 1440

ATCAACGGTA TTTAGGAGTA ATTCAATAAC CTGGGACTTT GTTAGTCGCA TTCCCCTTCT 1500

ATCTCTAGCA TCTTCTACTA AATTTTCAAG TTTCTCTAGA TTTTTATCAT CCAAGCTAAT 1560

CATTATTCTA TTTTTATCGG TTGCCATTTT CATCACCTCA AGTTAATTCT ATCACAGGTG 1620

TAACACTAGT GTCAACTGGC TTTTATAATA CATTAGTTTA AAAGTGGAGA GGATTTTTAA 1680

CACAGTAACT TTAAATCTTT GGTATTAAAA AATTTTCACA ATATTTATAG AAATAAAATC 1740

TGTCTCAAAT CAGTTATCAA ATCTAGTATA AATTATGAGC GGCTACTCTA ATACTTTCCC 1800

TCTAAACAAG AAAAAGACTT ACACTCAAGG GTTTTCTTCC CCCCCTTCGT TATAACGTTT 1860

TGACTCTTTT ACTAGCAAAG GTATATACTC ACAAGGAACT TTGGTTGACT ATTGAATCTC 1920

TCCAACTTCT TCTTTAACAT ATCCTTCTAC ATCTTCAATC TCTACAAACA TTGGGTCTAA 1980

GTGACACAAG AAATGCCAAA CTTCGATCCC TTTTTTTCTG TAAAGAATCG CTTCACCGTC 2040

TTCACTTCCG AAAAAGCTTC TGTCGATTTC ATATCCGCGG CTTTCTAAGA AGTCTTTTGC 2100

TTTACGATAG TTCGTTTCTC TTGTTTCGAC ATAGGCTTTA ACTTCATGGT TGTTAACGAC 2160

ATATGCATCA ATTTTTGAAT ATCCTTCGAT CACTCTATCA TTTTTGAGGG ATAAATTTGA 2220

AATCTCTTTC CAAATAATGT TTACATTTTC CTCAGGATCG AACATAAATT TAGATAAAGG 2280

AACAATATTT CCGTTAAAAA TAATTTCCAT ATAATCCGGT ATGTTTTTAG GATTAAAATA 2340

CTCCACTTCA AAACCATCTT CTGTTTCCAG AGTGTATCCC GGGATTTGAG CTACAAAGGC 2400

TTTCCCATCT TCTATGGAAT CAAATGCTAC TAAATCTTTA GAATAATCAT TTTGGTACAA 2460

TTCCAATATA ACCATCGATA ATCTCTCCAT TTTCATTATC AGGCTAATGT AAATAAGCAC 2520

GTCACCTGAC CAATTCAGGC TCTCTGTATC ATCTCATCAT ATTTCCTACT TACTTTACGA 2580

GTCTTATACC CAGAACACAC CTTATCGACC TTCGGTCTCA CCTCGTCGCA TTGGCTGAAC 2640

ATCTACTTTT ACTTTGCTGA TGCTTCAACT CGTACAAGCA GTGATACCGC CTCAGCGTGA 2700

TGCGTCAGTG GGACTCAAAA GGTTCGGGGA ACCTTTTGAG GATTAACTAC GTTTCTCTAA 2760

TAAACTTACA CATTCAACTT GTTCATCATT GTCCAAACCT ATGTTGAGAT TTTCTTCTAT 2820

AATTGGTAGC TTAAAAGTAA TGGATTTTAG CCATTGTCCG TTAGATTGTT TTTCTTCATA 2880

AACTTGAATT TCAGAAATCA AAGCTGAAAT TAACTGCCTA CGCTCTACAT CATTCATGAC 2940

TTTATAGAGC TTATCAAAAT AGATCAGAAC CTTATATATG TTATCTCCTG TAAGCTTTTC 3000

AGCTTCAATA GTCTGTTTCT TTGCTTTCGC ATCAATTAGT GATGATTCTA ATTCATCTAG 3060

TTTGTCATAC ATACGATATA GTCTATCATC TAAATCCTGT TTCCTTCTCT TATAATGCTT 3120

ATCTTCAACA TCTAAATTAT CTATTTCCTC AATTAGCTTA AACTTTGTAG AATGACTCTT 3180

TCTCAATTCC TTTTGGTAAT TATCTATTTC TTTTTCTATT TCAGAGGTAT CCACCTTCAT 3240

GTTGATTTTT TCTTGCATCA TAGAAGCAAA TTTCGGATTA CTTACTATCT TGACAATCAC 3300

CTCTGCAACA GCATCATCTA ACAATTCTTC TCTAATTTGC TTACTGAATG TACACTTATT 3360

ACCTCTTATC ATCTGCCTAT GGTTACAACC ATAGTAATAA AAATCTTTAT ACTTTGTGCC 3420

ATCTTTCTTT TTCTTGATAC ACTTGTTCCC AAACATTCCC ACTCCACATA TCGGGCATTT 3480

TACAATTCCA GAAAGCAAGT GTGTGCGTGT ATCTTTTCCT TTATTCACAT GCTCATATTT 3540

CTTTGCTTGA GATTTTAGCT TAACCTGAGC AGCTTGCCAA ACTTCATCGG AAACTATAGC 3600

TTCATGTATC CCTTCAGATA TTAGATATTC ATCTTGTTCA ACCTGCTTAT ATTCATTTCT 3660

TGTACCATGA ACTTTTTCTA AAGTTCTTCT TCCAAATGCT ATTTTCCCAT TATATACAGG 3720

ATTCTTTAAT ATCTTTCTTA TAAGACCTGC ATCAAACAAA GGATTCTTAC CATTCTGTCT 3780

TGGGATTTTT CTAATTCCAT GATTCTCTAA GTATTTAGAT ATCCCATTGG CTCCTATCGT 3840

AGTATTTACA TACTGGTCGA AAATCGTTCT TATTGCAACT GCCTCTTCCT CATTTATAAA 3900

CAGCTTGCCG TCTTCAAGTT TATATCCATA CGGAGCAAAG CCACCATTCC ATTTTCCTTC 3960

CCCTGCTTTT TGAATGCGAC CTTCCATTGT TTGAATACTG ATGTTTTCTC TTTCTATTTC 4020

AGCCACAGCT GATAAAACAG AAATCATTAG TTTCCCAGCA TCTTTAGATG AATCAATGCC 4080

ATCTTCAACG CAGATAAGAT TAACTCCATA ATCCTGCATT ATATGAAGTG TAGAAAGAAC 4140

ATCAGCGGCA TTTCTTGCAA ATCTTGATAA CTTAAACACA AGAACAAAAG ATACTCCATC 4200

TTTTCCAGAT TTTATATCTT CCATCATTCG ATTGAACTGT ATTCTACCTT CAATAGACTT 4260

GTCAGACTTC CCGGCATCTT CATACTCTCC AACAATTTCA TAATCGTTGT AAATAGCAAA 4320

AGCTTTCATT CGTGATTTTT GTGCCTCTAA CGAATACCCC TCTATCTGTA TTGACGTAGA 4380

TACTCGTGTA TAGAGGTATA CTTTTATTTT TTCTTTTGAC ATAGTATTAA CCTCAATATA 4440

ATTTTTCTAT ATCATATATA ATTTTTTTAA TTTAAGTTTG GACTATCATT TCAAGTATAT 4500

TATAACACTT TTATTAGTCC GTCTCAATTT GTGTTTTTGC CATGTCAAAA CTATTTTTCA 4560

TCTCTTGATT TTTTGCTGGC GTTGGATCGG GTAGATTATC TAAATCTAAA GCACCAGCAT 4620

ATTTTGCAAT CAGATTTGCT ATTAAATCAG CCAATCCATT CCAGTCATTG TCCAATATAT 4680

ACCTCCTCTA AAGTTTTATA TCTAATAATT ATTTGTTTAA TTAAGTTTTT TGACATTGAC 4740

AAGTGCTTTG GATTAGCAAC ATAGGAATCT CACTTCCGCC TCTATTCCGG ATGAGCCGGC 4800

TTCAACCTTA GAAGTATCAT TACCCTCATT TTCTTCATAG CGGATAGGGT ATCCCTCCCT 4860

ATATTCAAAC TCTTACTTAT CGCTCACTTT CTTTTTGCTT AGCAGAACTT TTTTTGCCGA 4920

ATTATTCAGC CGAAAGATCT TGACGGATAG GTTATTACGC TCCAAAAATA ATTAACGTCT 4980

TGTCTTGGTC TATTCAATTG TTAAGGTTCA AAATTTATCG AGAGTTATTA ATCTTTTTAA 5040

AATTTGACCA TCAGAAAATA TTTATCTTGA TGTAACAAAA TTCTATAAAT TACCCTCTTA 5100

TACTTAACAG TGAAAAGAAG TCTTTCTTGG TAACCAATTT TGAAATAGAA TTTGCTTATA 5160

TAAAAAGGTC CAATTCCCAC TGCATAAATA GCAGTGAAAA TTAGACCCTC TTGGTAACTG 5220

TCATCTAAAA GTCTTCTA 5238

(2) INFORMATION FOR SEQ ID NO: 151:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13425 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

GACGATTTAC GAAGAATCGA ACAAGAACCT GCTCCTATCA ATTCCCAACC TCTATCTCTA 60

AAATCTTGCA GTTCATGCTT ATACTTTTTT AAGAAATCTA GAATCATAGA TACGGTAGAT 120

GACATCGTCT GGTTGACATT GGTCAAAATA GAACAAACCA AAACGACTCG TTCTATACCT 180

CCAACCTTTC AAATGCATCT CATGTAAATG TTCTTCTTCC TTGTCCAAAT CAACAATGGT 240

GAAAATCCGA AATTCTACTC TGCTATTCAT TGTCTTACCC CAAAATTAGA AAACATGCCT 300

GGCGTTATTT ATTAGATAAT TCTTTCCACT TTTGACTCAA TCTCCAAAAA ATATAAGAAA 360

TCTGAATCGC AAAAACTATC AATAAAACCC AATCTATTAT GAAAATCAAA AACACTTTCC 420

AACTGAAAGA ACTACCTCCA GTGACAAACT TTGAGAAAAA CGGTAGTAGA GCTAAAAAGA 480

GAAATAAAAT AGGAAGCATC CGCATTGTTA AAATCCGTTT GGCATAAAAA AATCTTTATT 540

TAAACGAAAA TATTATGGCA AAATTTACGC CAGTTTTTGA ACGGCTGATG TAGATATTTT 600

ATACTTTCAA AATGTTTAAA TGTGATTATT TATTTTTGAA AAATAGATCA CCAGCCCGAC 660

TGAAAGTGCT TATAGAATGA TAATAAGTCG CCTGCCGAAA ACAGCGAAAA ATAGCGGTGT 720

TATGCGGAGA TAATCTGACG CGATGCGAAA GTATATTGCA TACTTATTTT CAACAATTTA 780

GCAGAGTATT TTTATAAGTG TGATATAATA GAAGTATAAT TTGTTCTGAT AGTTTATTTT 840

ATGGAGAAGT AGATTTTTAG AATGCGGAGG GTTCAATATG GTTGAGTTTA TAAAGTCTAA 900

GAAAGAAATG AGTGAGGAGG ATATTAAAGC AAATTTCATC ACTCCTGCTA TTGTATCCAA 960

1010

AGGATGGAAA AATGGTGAGC ATATCGCTTA CGAAGAATAC TTCACTGATG GTCGAATTGA 1020

AGTTAGAGGA GATAAGGCTC GTCGTAAAGA AGGAAAAAAA TCAGACTATT CACTGTATTA 1080

CCAATTTGGA ACTCGAATTG CAATTGTTGA GGCAAAGGAT AATAAACACA GCGTTCGAGC 1140

AGGATTACAA CAAGCTATTG AATATGGAGA GATTTTAGAT GTTCCATTTG TTTATTCTTC 1200

GAATGGTGAT GGCTTTATTG AACACGACCG TATCACGAGA GAAGAACGTG AGCTGGAGTT 1260

AGACGAATTC CCTACTCGTG AAGAATTATT TTCTCGTATG ACGAAGGAAA AAGGATTGAC 1320

GTACGAAATT ACAGAAGCTA TCTCAACTCC ATACTATACA GACGCCTTCT CAATGAAAAC 1380

GCCACGCTAT TATCAGCAAA TAGCTATCAA CCGTACTATT GAAACAGTTG CCAGAGGACA 1440

AAAACGAGTA ATGTTTGTGA TGGCAACAGG AACGGGGAAA ACGTTCATGG CTTTTCAAAT 1500

TATTCATCGC CTTCGAAAAG CTGGTTTGGC TAAACGAGTT TTATTCTTAG CAGATAGAAA 1560

CATCTTAGTA GACCAAACGA TGGCTGAAGA CTTTAGGCCA TTCGAAAAGG TAATGACGAA 1620

AATTACACCA AAACTTTTGA CTGCTCCTGA AAAATTAAAT TCTTTTGAAA TTTATCTAGG 1680

GCTTTATCAG CAACTAACTG GTGAAGATGG AACTGAAACA CATTATCAAA AATTTGACAA 1740

AGACTTCTTT GATTTAATCG TAATTGATGA AGCGCACCGT GGTTCAGCTA AGGAAAACAG 1800

TAACTGGCGT AAGGTAATTG ATTATTTCAG TTCTGCGACA CAGATTGGGA TGACCGCTAC 1860

TCTTAAAGAA ACCAAGAATG CTTCCAATAC GGAATACTTT GGTGAGCCAA TCTATACTTA 1920

TAGTTTAAAA CAGGGAATCG AGGATGGTTT TTTGGCTCCA TATCGTGTTA TGAGGGTTAA 1980

TTTAGATGTG GATGTGGATG GTTATCGTCC AGAAACTGGA AAAGTTGATG CTAACGGACA 2040

ATTAATAGAA GATAGGTACT ACGGCAGGAA AGATTTTGAT AAAACCATTG TCATTGATGA 2100

TAGAACGCAA AGAGTTGCCA AGTTTGTTTC TGATTATATG AAGCAAAACA ATGCACGATT 2160

TGATAAAACA ATTGTTTTTT GTGTTGATAT TGACCATGCC GAGCGAATGC GTGCTGCACT 2220

TGTAAAAGAG AATCTAGACT TAGTCCAAGA AGACTATCGT TATGTCATGC AAGTAACTGG 2280

TGACAACGCT GAAGGAAAAG CTCAACTGGA TAACTTTATG GATGTCAATT CTAATTTTCC 2340

CGCTATTGTA ACAACGTCTA AATTATTAAC GACAGGAGTT AATGCTAAAA CATGTCGTTT 2400

GATTGTTTTA GACTCTAATA TCCAATCCAT GACTGAATTT AAACAAATTA TTGGTCGTGG 2460

CACACGTCTT TATCCTCAAA AGGGGAAAGA ATTTTTTACG ATTATTGATT TTCGAAATGT 2520

TACCAATTTG TTTGCTGACC CTGATTTTGA TGGTGATCCA GTGAAGGTGC TAGAAACAGG 2580

TGCGAAAACA GTCAGTGGTT CTACGCCCGG TTTCGTAGAT GAGGAAGGTG ACCCAGTAGA 2640

AAAATATATC GTTACAGACA AGCAGGTTAC CATTCTTAAT TCTACTGTTC AAGTATTGGA 2700

TGAAAACGGG AAACTGATTA CCGAAAGCCT GACCGACTAC ACTCGAAAGA ATATCTTAGG 2760

TAGCTACGCC ACTTTGAACG ATTTTATCAC AGTTTGGCAT ACGGCAGATA AGAAGAAGCT 2820

TATCTTAGAC GAACTTTATA AAAAAGGAGT TTATCTAGAT GCTATTCGAG AGTCGGAGGG 2880

AATATCAGAA CAAGAAATCG ATGATTTTGA TTTACTCCTA AAACTTGCCT ATGGTCAAAA 2940

AGAATTAACC AAAACGGAAC GTATCAATAA ACTCAAACAA AGCGGATATT TATATAAATA 3000

TAGTGAGGAA GCGCGTGCTG TTTTGGAAAT TTTACTGAAC AAATACATGG ATAAAGGTAT 3060

TGGAGAACTC GAAAGCATTG AAACATTAAA ACTTCCAGAA TTTCAGATAT ATGGTGGAAC 3120

CTTCAAAATC ATCAATACTT ATTTTGGAGA TAAAAAACGA TATTTACAAG CAATTAAAGA 3180

ATTGGAGCAA GAGCTATTTA CAGTAGCTTA ATGAAAGGAA AGTATGTCAA TTACATCATT 3240

TGTAAAAAGA ATTCAAGATA TCACTCGAAA CGATGCTGGT GTTAATGGTG ATGCTCAACG 3300

TATTGAGCAA ATGTCTTGGT TATTATTCTT AAAAATTTAT GATAGCCGTG AAATGGTTTG 3360

GGAATTAGAA GAAGACGAGT ATGAGTCAAT TATCCCAGAG GAATTAAAAT GGCGAAATTG 3420

GGCTCATGCT CAAAATGGGG AACGGGTATT GACAGGCGAT GAATTACTTG ATTTTGTCAA 3480

TAACAAGTTA TTCAAAGAGT TGAAAGAGCT TGAAATAACT TCAAATATGC CTATTCGAAA 3540

AACGATTGTT AAATCAGCTT TTGAAGATGC GAACAACTAT ATGAAAAATG GCGTCTTGTT 3600

ACGCCAAGTC ATCAATGTTA TTGATGAAGT TGATTTCAAT AGCCCTGAAG ATCGTCATTC 3660

GTTTAATGAT ATTTACGAAA AAATTCTTAA AGATATTCAA AATGCTGGGA ACTCAGGAGA 3720

ATTTTATACG CCACGTGCAG CGACTGATTT TATTGCCGAA GTTCTTGACC CAAAACTTGG 3780

AGAATCAATG GCAGACCTTG CTTGCGGAAC AGGAGGCTTC TTGACTTCGA CTCTGAACCG 3840

TTTAAGTAGT CAACGTAAAA CTAGTGAAGA TACCAAAAAA TATAATACAG CTGTTTTTGG 3900

TATTGAAAAG AAAGCATTTC CTCATCTTTT AGCAGTTACA AATCTGTTTC TTCACGAAAT 3960

TGATGACCCT AAAATTGTTC ATGGAAATAC TTTGGAGAAA AATGTTCGTG AATATACGGA 4020

TGATGAAAAA TTTGACATTA TTATGATGAA TCCACCTTTT GGAGGGTCAG AATTAGAAAC 4080

AATAAAAAAT AACTTTCCAG CAGAATTACG GAGTTCTGAA ACAGCTGATT TATTTATGGC 4140

TGTCATTATG TATCGTTTGA AAGAAAATGG TCGTGTTGGA GTTATTTTAC CTGATGGTTT 4200

TCTATTTGGT GAAGGTGTAA AAACTCGCTT GAAACAAAAA CTGGTAGATG AGTTCAACTT 4260

GCATACGATT ATTAGGTTGC CTCATAGTGT CTTTGCACCG TATACAGGAA TCCATACGAA 4320

CATTCTTTTC TTTGATAAAA CAAAGAAAAC AGAAGAAACT TGGTTTTATC GTTTAGATAT 4380

GCCAGATGGT TATAAAAATT TCTCGAAAAC TAAGCCGATG AAGTCAGAAC ACTTCAATCC 4440

TGTTCGTGAC TGGTGGGAAA ATCGTGAAGA GATTCTGGAA GGTAAGTTCT ACAAATCTAA 4500

ATCATTTACA CCTAGTGAAT TGGCTGAGTT GAATTATAAT TTAGACCAGT GTGACTTTCC 4560

AAAAGAGGAA GAGGAAATCT TAAATCCCTT TGAGTTGATT CAGAATTATC AAGCGGAAAG 4620

AGCAACTTTA AATCATAAGA TTGATAATGT ATTAGCTGAT ATTTTGCAGT TGTTGGAGGA 4680

CAAATAATGA CACCAGAACA ACTTAAAGCA AGTATTCTCC AAAGAGCGAT GGAAGGGAAA 4740

TTAGTGCCGC AAAATCCCAA TGACGAACCT GCAAGTGAAT TATTAAAGAG AATTAAAGCT 4800

GAAAAAGAAA AACTTATCAG TGAAGGAAAA ATCAAACGAG ATAAAAAAGGA AACTGAGATA 4860

TTTCGTGGTG ATGATGGGAA ACATTATGGG AAGTTTGCTG ATGGAAGCAC TCAAGAAATT 4920

GATGTTCCTT ATGATATTCC TGATACTTGG GAGTGGGTGA GGTTTTCTAC ATTGGTTGAA 4980

ATTGTCAGAG GTGGCTCTCC ACGACCAATC AAaGATTATC TTACTTCTGA AGTAGATGGA 5040

ATAAATTGGA TAAAAATAGG TGATACTGAA AAGGGTGAAA AGTATATAAA TAATGTTAAA 5100

GAAAAAATCA AAAAATCAGG GCTTAACAAA ACTAGATTTG TAAAAAAAGG TACATTTTTG 5160

TTAACTAATT CTATGAGTTT TGGTAGACCT TATATTTTGA ATGTTGATGG TGCAATACAC 5220

GATGGATGGT TGGCTATTTC GAACTATGAA AACTCATTAA ATAAAGATTA CCTATTCTAT 5280

ATTCTTTCAT CAAATGTAGT TTATTCTCAA TTTCTATCTC TAATTAGTGG AGCTGTTGTG 5340

AAAAACTTGA ATAGTGATAA AGTTGCTTCT ATTCTTATCC CTCTCCCCCC ACTATCCGAA 5400

CAACAACGAA TAGTAGAAGC AATCGAATCA GCTTTAGAAA AAGTAGATGA ATATGCTGAA 5460

AGTTATAATA GACTAGAACA GCTAGATAAA GAATTTCCAG ATAAACTAAA AAAATCTATT 5520

CTTCAATATG CTATGCAAGG AAAATTAGTT GAACAAGACC CAAATGATGA ATCAGTCGAA 5580

GTTTTACTTG AAAAAATACG AGCAGAAAAA CAAAAACTCT TTGAAGAAGG CAAGATTAAA 5640

AAGAAAGATT TGGACATTTC TATTGTTTCC CAAGGAGATG ATAACTCTTA TTATGGGAAT 5700

ATACCTATGA ATTGGGTTGT TATAAAAATA AAAGATATTT TTTCAATAAA TACAGGTCTT 5760

TCTTACAAGA AGGGCGATTT AAGCATTAAT AATAAAGGTG TTAGAATTAT ACGTGGTGGT 5820

AATATTAAGC CTTTAGAATT TTCTCTGTTG GATAATGATT ACTACATTGA TACACAATTC 5880

ATCTCCTCTG AGCAAGTTTA TTTAAAACAT AATCAGCTAA TAACACCTGT ATCAACCTCT 5940

TTAGAACATA TTGGAAAGTT TGCAAGAATC GATAAAGACT ATGATGGTGT TGTGGCTGGT 6000

GGATTTATTT TCCAATTAAC ACCATTCGAA AGTTCAGAGA TTATTTCAAA ATTTCTATTA 6060

TTTAACTTGT CCTCTCCGTT ATTTTATAAA CAATTGAAAG CAATAACTAA ACTATCAGGT 6120

CAAGCTTTAT ATAATATTCC TAAAACTACA CTGAGCGAGC TATTAATTCC GTTAGCTCCT 6180

TTTGAGGAAC AGGAACTTAT TACTCAAAAA GTTGAGAAAC TTTTTGAAAA AGTAAATCAA 6240

CTTTGAAAAT GATTCTTTTC ATCTCTTCAT GATTAGAAAT AGGGATTAAT AATTCGGAGA 6300

TACTGGTACT ATTTAATGTT TTCCCTTTGA TAGCATCTTT TGAATCACCT AAAGTAGAGA 6360

TAAGTGGCAA AAATATCATT AAGTAATCTC TGATAATATT TTCTTTATTA GCATAGGGGA 6420

ATATCGATAT AATGGCTTCA TTATGAGTGG CAGGAATATC CAATATGGCA ACTTTTCCAA 6480

TAGATAATTT AAAACTCATT AATAAAGTTC CTTTAGGTGA AATGTCTATT TTCTTTGATT 6540

TTAATGCTAA TTTAGAAATA GATTCTCTCG CATTAGTTAC ATAACCAGAT ATAGGCATAT 6600

CTGATATAGA TACCCAAGGT ATTTCAGTTC CCCAAAAAGT AGCTTCACTG CGTGGAGGAG 6660

TTTTTCCTAT TCTGAAGTTA ACTAGGCTAG CAAATTTAAT ATATCTCCAT GCTTCTGGGA 6720

TTTCATATAT AGGATAAGAG GTTGTTTCGT CTTTGTTCCC ATAATAAGAG CCATAATCAC 6780

AAAAATAGCA GGTAGTCAGT TTGACCACCT GTTATTTTTT ACCAATTAAC AATTTTATCT 6840

ACAATATTTT GTTGTTCAGT AGCTGTTTTC CTTAGATAAA TTCGAGTAGT TTCTATACTT 6900

TCGTGTCCCA TCAAATCTGC AAGCAAGGCA ATATCATTAT ACTTCGCTAA AAAATTCTTA 6960

GCAAATAAAT GCCTAAAAGA ATGAGGGTAA ATTACGTTAG GATTCATTTT GTATTTATCA 7020

GCATAATTTT TTAACTGTTG AGCAACTCCT CTTGCTGTAA TTGGTTCGTT AAATTTATTC 7080

AAAAATAAAT AACCACTTCG GCGATTTTCT GATTCTAACC AACTAAGACA ACTATTTCTT 7140

AATTTTTTAG GAATGTACAG TCTACGAATT TTACCACCTT TTGAGTAAAT GTCAAAATAA 7200

CCGATTTCTA CATGCTCTAC TTTTAGTTTA ATAAGTTCAC TTACACGAGC CCCAGTTGCA 7260

CCTAAAAACC AAACGACAAA ATGCCATTTT AAAATACCAT CTTTTTTCAA ACTACGTTTA 7320

AGAAAAAGGT AATCAGCATG GCTAATGACA TCTTCTAAAA ACGGTTTTTG CTGTACTTTG 7380

ACAAATTTTA ATTTCAAATC ATCATGACCA ATAAAAGCCA GATATTTATT TACTCCTTGT 7440

AGTCGCAAAT TGACAGTTTT AGGTTTAAAA TTGTCTAATA AATATCCTTT GTATTCAAAT 7500

AAATCTTCCA TTTTGAGTTC GTAATTCTCC AAGAAAAATC GAACACCATA AAGGTACGAA 7560

CGCACAGTAT TTTCAGCTAA ACCAGCTTTC TTCAAATGTA ATTCAAAATC TTTCAACGTA 7620

AAACTCCTAT CTTATGTTTG ATAGAAATTC CACCGCACGT AAAACTATTA TACTAAATTA 7680

GTGCGTCAAT ATGGGCGAAA AATTGTTCGA TTTTATCAAC GATTCTGGAT TGTTCAGGAA 7740

GGGGTGGGAG GGGGATTAAA TATTCTTTTA TAGTTTTCGT TAATAATTCT TTTTGTTTTG 7800

TACTACCCGA CGCTTTTTCT TCAATAACTG ACTGAACAAT AGGAGAGGAA AGAAAATTAT 7860

AGATGAAATG GCAATTAATA ACCCCCGATA AGACTCTTAT AACTGTAACA TGGCTATCTG 7920

CAACAGCCCA GCCATAAGGA TTTTTATTTT CATGGTAAAT AGCTAATCGT CCTAACGTAC 7980

CTAGACCTGT TGAATTCCAC ATTAAATCAC CATCTCTTAG TAATCTTTCT TTCTGGTAAC 8040

TATGAACTGT TTCGGGATCA ATAAATCTTG CTAAGTCAAT AGAAAAGCCA GACCATTGAT 8100

TACATTTCTG AGCAATCACA GGGTATATAG GAATATTTGA ATATTTTGGA GACTTCCCTC 8160

TTTGAATGTA GGAGGTTATA TCGTTTAACC TCACCCATTC CCAACTTTCT GGTATTTCAC 8220

AAGGTACTTC CTCATAATAA GAGTTATCAT CTCCTTGGGA AACAATAGAA ATGTCCAAAT 8280

CTTTCTTTTT AATCTTGCCT TCTTCAAAGA GTTTTTGTTT TTCTGCTCGT ATTTTTTCAA 8340

GTAAAACTTC GACTGATTCA TCATTTGGGT CTTGTTCAAC TAATTTTCCT TGCATAGCAT 8400

ATTGAAGAAT AGATTTTTTT AGTTTATCTG GAAATTCTTT ATCTAGCTGT TCTAGTCTAT 8460

TATAACTTTC AGCATATTCA TCTACTTTTT CTAAAGCTGA TTCGATTGCT TCTACTATTC 8520

GTTGTTGTTC GGATAGTGGG GGGAGAGCAA TTAATAATAG ATTAAAATTA TAATCATTGA 8580

TTGCAGGATA ACTTGTTCCA GTAGATTTAT TATTAACACG ATTGATAAAA TTATCTGATA 8640

ATAAATAATA TTTCAAATAT GTTTCGTTAA GTAAAGTATC CAAAACAATA AATGCTGTAC 8700

TAGCTATCAA ATACTCTTTA AGTTCTCTAA CTACAGCAAT ATTTTTTAGA TATGGTCTAA 8760

CTGTTGAAAA TAAGACACTA TTCTGCGAAA CTAATTTTCT AGCACGGGAA GGCGCTTGTT 8820

CAGGTGAAAG ATATTGTAGA TTTTTGTAGT TGATTATGTT CTTTTTTCTA TCAATACTAG 8880

ACGTATCTAT ATACCTAAAG GATTTCTCTG GCTTATTTTG CCCAAAATTC CAATAAATTG 8940

ATTTTATCCT CACCCACTCC CAAGTATCAG GAATATCATA AGGAACATCA ATTTCTTGAG 9000

TGCTTCCATC AGCAAACTTC CCATAATGTT TCTTATGTGC TTCAAGTATA TAAAAAGGCG 9060

TAAAAATACG CCTATAGATA ATGGGGTTGA AATAGGTTTA TTGTTGATGA GATTGTAGAT 9120

AATTCAATTT TTTACTTCCA ATCGAATATT CAAATCCTCC ACCTTTTCTG CCTGTAATTG 9180

TTCATCATAA AATTCAATAT CTTCAGGATT TTCCCCTTGG CAACCTCGGC AGAAATATTC 9240

TTCCGCTCGA TCAGGATTCA AAAATCGACA AGCACAAACA AAACAGTCGC CATCATCATT 9300

TATTGAGATA ATATAGTAGA TTGAAATAAG ATGTAAACAA ATCGATTAGG AAAGTTAAAT 9360

TAGTTTCTAG AAATTTTTAG CAGATGTAGT GTACTATTCT AGTCTCAATT TACTATGGCT 9420

TCAAATATAT CTTTCGAAAA AATATTTACA GATGTGTAAT TTTGAAGCTT GCAAAAGTTA 9480

GTAAACTTGT AGATTTCGAT TTGAAGTAAC TTGTTTTCTT GCCCGATATT GTTTTTGAAA 9540

TTGAATTTTT CCATAGTGAC TCCTTAATTT TCTTCTACAC GTCTGATGAT AAATCTAATT 9600

CGCAAAAGAG TCAAGAGGAT TTTTCGAAAA ATAAATAGCG ACCGAAATCG CTATTTTAAG 9660

GGTTATAGGT ATTTGATGGC TTAGACTGCT GTGTGACTGT TTACCCACAG GCAATCTTTC 9720

TTCTATATTA GTATTAGTAA AGGTCTAAAT AATTATCAAT TTCCCATTGT GAAACGAAGG 9780

TTGCATAACT TGCCCATTCG ATTCGTTTGG CTTCAAGGAA GCTAGTATAG ATGTGATCTC 9840

CGAGAGCAGC TTTAACCACT TCATCTTCTG TCAAAGCTTT CAAAGCGTTG TGAAGAGTTG 9900

ATGGAAGGTC TGTAATACCA GCTTCCTTGC GCTCTTCTGC TGTCATGATG TAGATATTTT 9960

CTTCGATAGG AGCTGGTGCT TCGATTTTAT TTTCAATACC ATACAAACCA ACTTCCAAAA 10020

GAACAGCCAT AGCAACGTAA GGGTTCGCCA TTGGATCCAC TGAACGCAAC TCAAGACGAG 10080

TTCCCATACC ACGTGAAGCA GGTACGCGCA CAAGTGGCGA ACGGTTACGA CCAGCCCAAG 10140

CAATGTAAAC AGGCGCTTCA TAACCTGGAA CCAAACGTTT GTATGAGTTA ACTGTTGGGT 10200

TCATGATGGC AGTATAGTTG TAAGCATGCT TGATCAAACC GCCTAGGAAA TGGTAAGCTG 10260

TTTCTGACAA CTGCATTCCT TTTGGATCAT TTGGATCAAA GAAGGCGTTA TTTCCTTCTG 10320

CATCAAACAA GGACATATTA CAGTGCATAC CTGATCCAGC AATACCAAAT TTTGGCTTCG 10380

CCATAAATGT TGCGTAAAGT CCGTGTTTGC GAGCAATGGT TTTAACAACA AGCTTAAAGA 10440

TTTGAATCTT ATCACAAGCA CGGAGAACTT CATCGTACTT AAAGTCAATC TCATGCTGTC 10500

CAACCGCAAC CTCGTGGTGA CTCGCTTCTA CTTCAAATCC CATTTTGGTC AAGACATTCA 10560

CAATCTCACG ACGTGTGTTG TCCGCAAGGT CAGTAGGTGC CAAGTCAAAG TAGCCACCCT 10620

TGTCATTCAC TTCAAGTGTT GGGTCCCCAT TTTCATCCAA CTTAAATAGG AAGAATTCTG 10680

GCTCTGGACC AAGGTTGAAG GATTTGAATC CAACTTCTTC CATGTGACGA AGAGCTCGTT 10740

TCAAATTACC ACGAGGGTCA CCCGCAAATG GTTCACCTTC TGTTGTATAG ACATCACAGA 10800

TCAGACCTGC AACACTTCCA TTTTCATCTC CCCAAGGGAA GACTGTCCAT GTATCCAAGT 10860

CCGGGTACAA GTACATATCC GACTCATTGA TACGTACAAA ACCTTCAATA GAAGATCCAT 10920

CAAACATAAC CTTGTTCGAC AAGACCTTAT CTAACTGTTC ATCTGTAGCA GGAATTTCGA 10980

CGTTTTTCAT GGTTCCCAAA ATATCTGAGA ACATAAGACG AATAAAGGTA ACATTTTTTT 11040

CCTTGACTTC ACGACGAATA TCTGCAGCTG TGATTGGCAT AAGTTTTCTC CTTAATCTAT 11100

GACTACTTGC GGTTGCCTAA CCGCGACCAA AAGGTGACTG TACTGAAGCA AAACGCCCCT 11160

GTTGGAGGAG TTCATTGTGA AGTGCACGAC GTACTTCAGT CTGACTAACC GCTTTCTTGG 11220

ATTTCGCTTC ACGTTCAGCA TATTTTTTCT TAATGGCAGC GATATTATAA CCTTCAGAGA 11280

TATAATCTTT GATTTCAAGC AGACGATCCA TGTCATTCAA GGAATACATG CGACGATTTC 11340

CTTCGTTTCG ATCGGGCTTG ATCAACTCTT GATCTTCATA ATAACGAATC TGACGCGCCG 11400

ATAGATCGGT CAACTTCATA ACACTGCCGA TAGGAAAAAC AGCCATATTT CGGCGAAATT 11460

CTTTTTCCTT CATTTACAAT TTCCTTCTTT CTGTCTATTA TAGTCTAAAA AAAGACAAAC 11520

GTCAATTGAT AATGTTATAA AATGTAACAT TATTTTTCTT TTTTCTCTAA AAAGAGACGA 11580

ATACGATCAA TATCGTAATT TACGATAATT GCGACAAAAA CTCCCATAAA CGTTTCTAAT 11640

ACACGCACAA ACACGTACAA AATTGTCTCA CCACTTGGAA TTGATAGGGT AATGATTAAC 11700

ATAGCTGCTA CACCACCAAT AACCCCTGCT TTGTTATTCA TGGCTACATT TGTCATAATG 11760

GTTAACATGG TGCAGATTGG AACAACTACC AAGGTCACCC AAAAGGCTTC GTGGAAAAAG 11820

GTATTTAATA AGAAGAAGAC CAAGGCATAG AGTCCACCGA TACTATTTCC TAGAATACGC 11880

GAAGTCCCAA AATGAACACT CTCATCAAAA CTCTCCCTCA GGCTAAAAAC GGCTGTCAAA 11940

GCACCAATTT GAAGACCTTT CCAGCCAAAA AAGCCAAAAA TCAAGAGAAC TAGAAAAACA 12000

GCAATACCTG TTTTAAAGGT TCGCATACCA AGTTTGAACT GGGATTTATC GAATTTATAT 12060

TTTTTAAAAT AACTCATAAT CTCAACTTTC TATTTCCATT TTATCATAAA TCGGTGATTT 12120

TTATGAGTAA TAGTTGAGAG GAAGCGTTTT TATTTTAAGC AAAAGAAAAG AGGAACTTTC 12180

ATCCCTCTCT TCTTTGATTT ATTTATAAAA TCTTATTTTT CTGTCAAGGC TGCAAGTCCT 12240

GGAAGAACCT TACCTTCAAG AAGTTCCATT GATGCTCCAC CACCCGTACT AATCCATGAG 12300

AACTTGTCTG CACGGCCAAG GTTAATCGCT GCGGCAGCTG AGTCACCACC ACCGATGATT 12360

GATTTAACTC CTGGTTGTTT CACGATAGCG TCCATCACAC CGATTGTACC AGCTTGGAAA 12420

TCTGGGTTTT CAAATACACC CATAGGTCCG TTCCATACGA CTGTTTTGGC ACCAGTCAAA 12480

GCTTCGTCAA ATTTGGCGAT AGATTTTGGA CCGATGTCAA GACCAAGGAA GCCTTCAGAA 12540

ACTGCTTCAC CTTCAGTGTC ACGCACTTCA GTGTAACCAG CAAATGCGTT AGCTTCTTTT 12600

GAGTCAACTG GCAAGATCAA TTTACCATTT GCTTTTTCAA GAAGAGCTTT CGCAACATCC 12660

AATTTGTCTT CTTCTACAAG TGAGTTACCG ATTTCGATAC CTTGTGCTTT GTAGAATGTG 12720

TAAGTCATCC CACCACCGAT AAGGACGTTA TCAGCTTTTT CAAGCAAGTT TTCGATAACA 12780

CCGATCTTGT CTGAAACTTT TGAACCACCA AGGATAGCCA CGAATGGACG TTCTGGAGTT 12840

TCAACTGCTT CTTGGATGTA GGCAATTTCG TTTTCAAGAA GGAAACCAGC AACTGCTTTT 12900

TCAACGTTTG CTGAGATACC AACGTTAGAT GCGTGTGCAC GGTGAGCTGT ACCGAATGCA 12960

TCGTTTACGA AGATACCATC TCCAAGTGAT GCCCAGTATT TACCAAGTTC AGGATCGTTT 13020

TTAGATTCTT TCTTGCCGTC AACATCTTCG TAACGAGTGT TTTCAACCAA GAGAACTTGT 13080

CCATCTTCAA GAGCGTTGAT TGCCGCTTCT AATTCAGCAC CACGAGTGAC ACCTGGGAAA 13140

ACAACATCTT GACCAAGTTT TGCTGCCAAG TCAGCTGCTA CAGGAGCAAG TGATTTACCA 13200

GCTTTATCAG CTTCTTCTTT CACACGTCCA AGGTGAGAGA AAAGAATTGC ACGTCCACCT 13260

TGTTCGATGA TGTACTTAAT AGTTGGAAGA GCTGCTGTGA TACGGTTATC GTTAGTGATT 13320

ACGCCATCTT TCAATGGTAC GTTGAAGTCA ACACGAACGA GGACTTTTTT ACCTTTCAAG 13380

TCAACGTCTT TAACAGTAAG TTTTGCCATG TTACAAAAAC TCCGG 13425

(2) INFORMATION FOR SEQ ID NO: 152:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 905 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

GATTTATCCT ACCGGnGAAT TTCCGGAGGG GTTCTAGCAG CAATCTTAGG AATCTATGAA 60

CGAATGATTG GCTTTCTGGC CCATCCCTTT AAAGACTTTA AAGAAAATGT TTTGTACTTT 120

ATTCCAGTTG CCATCGGTAT GCTTCTGGGA ATCGGCTTAT TTTCCTACCC GATTGAATAC 180

CTGCTTGAAA ATTATCAGGT TTTTGTATTA TGGAGCTTTG CGGGAGCTAT TATCGGTACA 240

GTTCCTAGCC TCCTCAAAGA ATCAACTCGA GAATCTGACC GAGACAAGAT TGATTTAGCT 300

TGGTTATGGA CAACCTTTAT CATTTCTGGA TTAGGACTCT ATGCCTTAAA TTTTGTCGTT 360

GGAACCTTAA GCGCCAGCTT TCTTAACTTC GTCCTAGCAG GCGCACTATT GGCCCTTGGC 420

GTCTTGGTTC CTGGCCTCAG CCCATCAAAT TTACTTTTGA TTTTGGGACT CTATGCTCCT 480

ATGTTGACTG GTTTTAAAAC TTTTGATTTC TTGGGAACCT TCTTTCCGAT TGGAATTGGT 540

GCAGGTGCAA CTCTCATCGT TTTTTCAAAA TTGATAGATT ATGCCTTAAA CAACTACCAC 600

TCACGCGTCT ATCATTTCAT CATCGGTATC GTCCTATCAA GTACCCTTTT GATCTTAATT 660

CCAAATGCAG GAAACGCTGA AAGTATCCAA TACACAGGAC TTTCACTTGT CGGTTATGTC 720

ATCATCGCCT TCTTCTTTGC GCTGGGAATC TGGCTTGGTA TTTGGATGAG TCAATTGGAG 780

GATAAATATA AATAATGGCA AAAAAAGTTA AAATCAAAAA AACATTGGTG GAACAAATCC 840

TATCTAAAGC AGCTATCCCT CATCAGGGGA TTCAAATCAA TGCCCTAGAA GGAGAGCTTC 900

CTCAA 905

(2) INFORMATION FOR SEQ ID NO: 153:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4278 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

1018

CTTGAATTAA ATAAAAAACG TCATGCGACT AAGCATTTTA CTGATAAGCT TGTTGATCCC 60

AAAGATGTGC GTACGGCTAT CGAAATTGCA ACCTTAGCGC CAAGCGCCCA CAACAGCCAG 120

CCTTGGAAAT TTGTGGTGGT ACGTGAGAAA AATGCTGAAC TGGCAAAGTT AGCTTATGGT 180

TCCAATTTTG AACAGGTATC ATCAGCGCCT GTAACCATTG CCTTGTTTAC AGATACGGAC 240

TTAGCCAAAC GTGCTCGTAA GATTGCCCGT GTTGGTGGTG CTAATAACTT TTCTGAAGAG 300

CAACTTCAAT ATTTTATGAA AAATCTGCCA GCTGAGTTTG CCCGTTACAG TGAGCAACAA 360

GTCAGCGACT ACCTAGCTCT CAATGCAGGT TTGGTTGCCA TGAACTTGGT TCTTGCATTG 420

ACAGACCAAG GAATTGGTTC TAACATTATT CTTGGTTTTG ACAAATCAAA AGTTAATGAA 480

GTTTTGGAAA TCGAAGACCG TTTCCGCCCA GAACTCTTGA TCACAGTGGG TTATACAGAC 540

GAAAAATTGG AACCAAGCTA CCGCTTGCCA GTAGATGAAA TCATCGAGAA AAGATAGAAA 600

GAAGAAAAAA TGACAGCAAT TGATTTTACA GCAGAAGTAG AAAAACGCAA AGAAGACCTC 660

TTGGCTGACT TGTTTAGCCT TTTGGAAATC AATTCAGAAC GTGATGACAG CAAGGCTGAT 720

GCCCAGCATC CATTTGGGCC TGGTCCAGTA AAAGCCTTGG AGAAATTCCT TGAAATCGCA 780

GACCGCGATG GCTACCCAAC TAAGAATGTT GATAACTATG CAGGACATTT TGAGTTTGGT 840

GATGGAGAAG AAGTTCTCGG AATCTTTGCC CATATGGATG TGGTGCCTGC TGGTAGCGGT 900

TGGGACACAG ACCCTTACAC ACCAACTATC AAAGATGGTC GCCTTTATGC GCGCGGGGCT 960

TCGGACGATA AGGGTCCTAC AACAGCTTGT TACTATGGTT TGAAAATCAT CAAAGAATTG 1020

GGTCTTCCAA CTTCTAAGAA AGTTCGCTTC ATCGTTGGAA CAGACGAAGA ATCAGGCTGG 1080

GCAGACATGG ACTACTACTT TGAGCACGTA GGACTTGCCA AACCAGATTT CGGTTTCTCA 1140

CCAGATGCTG AATTTCCAAT CATCAATGGT GAAAAAGGAA ATATCACGGA ATACCTCCAC 1200

TTTGCAGGAG AAAATACAGG TGTTGCCCGT CTTCACAGCT TTACAGGTGG TTTACGTGAA 1260

AATATGGTAC CAGAATCAGC AACAGCAGTC GTTTCAGGTG ACTTGGCTGA CTTGCAAGCT 1320

AAACTAGATG CCTTTGTTGC AGAACACAAA CTTAGAGGAG AACTCCAAGA AGAAGCTGGC 1380

AAATACAAGG TGACGATCAT TGGTAAATCA GCCCACGGTG CTATGCCTGC TTCAGGTGTC 1440

AATGGCGCAA CTTACCTTGC CCTCTTCCTC AGCCAGTTTG GCTTTGCTGG TCCAGCCAAA 1500

GACTACCTTG ACATCGCAGG TAAAATTCTC TTGAACGATC ATGAGGGTGA AAATCTTAAG 1560

ATTGCTCATG TGGATGAAAA GATGGGTGCT CTTTCTATGA ATGCCGGCGT CTTCCACTTC 1620

GATGAAACAA GTGCTGATAA TACCATTGCC CTCAACATCC GCTATCCAAA AGGAACAAGT 1680

CCAGAACAAA TCAAGTCAAT CCTTGAAAAC TTGCCAGTTG TTTCTGTTAG CCTGTCTGAA 1740

CACGGTCACA CGCCTCACTA TGTGCCAATG GAAGATCCAC TTGTGCAAAC CTTGTTGAAT 1800

ATCTATGAAA AACAAACTGG CTTTAAAGGT CATGAACAAG TCATCGGTGG TGGAACCTTT 1860

GGTCGCTTGC TAGAACGCGG AGTTGCCTAC GGTGCTATGT TCCCAGACTC GATTGATACC 1920

ATGCACCAAG CCAATGAATT TATCGCCTTG GATGATCTTT TCCGAGCAGC AGCAATTTAT 1980

GCCGAAGCTA TTTACGAATT GATCAAATAA AACGATAGAA GTCTGAGATC TTATGCTTGG 2040

ACTTCTTTTT GGAGGGAAAG TAGATGTCTC AAATCGAAAG AATCAAACAG GCTATCATGG 2100

CGGATTCGCA GAATGCCAGC TATACAGAGC GTGGCATTGA GCCTCTCTTT GCAGCGCCAA 2160

AAACTGCTCG CATCAATATC ATCGGTCAGG CTCCGGGACT TAAAACTCAA GAAGCAGGCC 2220

TTTACTGGAA AGATAAAAGT GGTGACCGCT TGCGGGACTG GCTAGGTGTG GATGAAGATA 2280

CCTTTTACAA TTCAGGTTAT TTTGCTGTTT TGCCTATGGA TTTCTACTTT CCAGGACATC 2340

GCAAGTCGGG TGATCTTCCG CCTCGTACAG GTTTTGCAGA AAAATGGCAT CCGCAGGTCT 2400

TACAGGAATT GCCTGATATT CAGTTAACCC TCTTGATTGG GCAATATGCC CAAGCCTACT 2460

ATTTACAGGA GAAAATCAGT GGGAAGGTAA CGGAGAGGGT GAAACACTAT AAAGACTATC 2520

TGCCAGCCTA TTTTCCGCTA GTTCACCCAT CACCACGAAA TCAAATCTGG ATGGCCAAAA 2580

ATCCTTGGTT TGAGGCAGAA GTAGTGCCAG ATTTGAAAAA AAGAATTAAA ACCATTTTAT 2640

AGTCAATGAA AATCAAAGAG CAAACTAGGA AGCTAGTCGT AGGCTGCTCA AAGTACAGCT 2700

TTGAAGTTGC AGATAAAACT GACGAAGTCG GTAACATACG CACGGTAAGG CGACGCTGAC 2760

GTGGTTTGAA GAGATTTTCG AAGAGTATTA GAAGAAAAAG AATGAAAGAA ATAGCCTTTG 2820

ACGCATTTTA CCAGCTTTAC CAAAACGACC AGCTTTCTTT AGTGGATGTG AGAGAAGTGG 2880

ATGAGTTTGC AGCTCTTCAT TTAGAAGGTG CCCACAACCT ACCGCTTAGT CAATTGGCTG 2940

ATAGTTATGA TTAATTGGAC AAAGATCGCT TGCATTATAT TATTTGCAAA TCTGGAATGA 3000

GATCGGCGCG TGCTTGCCAA TTCCTATTAG AACAAGGTTA TAATGTTATC AATGTCCAGG 3060

GTGGCATGTT AGCCTTTGAA GAACTTTAAA ATTTTGCATT TCTCCTACTT GGTGTGGACT 3120

GGGTAGGAGA GTTTTATTTT TAGATAATTC TTATTTTTAA GAAAATTGAA AACATTTAAT 3180

ATTTGCCTCG TGATGCTTTT TTCAGACTCC TAATCGTGGT ATACTAGGTC AGTATTTTAT 3240

AAATATGAAG GAGATTTTTA TGGCTAAAAA AGGTACCCTA ACAGGTTTGC TCCTGTTTGG 3300

AATATTTTTT GGTGCGGGGA ACTTGATTTT TCCGCCTTCT CTAGGTGCTC TATCTGGAGA 3360

ACATTTTCTT CCTGCCATCG CAGGTTTTGT CTTTTCAGGC GTTGGTATCG CCGTCTTGAC 3420

CCTTATTATT GGAACGCTAA ATCCTAAAGG ATATATCTAC GAGATTTCAA CGAAGATAGC 3480

GCCTTGGTTT GCGACTCTTT ACCTCTCAGT TCTTTACTTG TCAATCGGTC CATTCTTTGC 3540

1020

TACCCCACGT ACTGCTACAA CAGCTTACGA AGTAGGGATT AGCCCCCTTT TGTCGGATGC 3600

AAATAAAGGA CTTGGCTTGA TTGTATTTAC GGTTCTGTAT TTTGCGGCAG CCTATTTGAT 3660

TTCGCTTAAT CCATCAAAAA TCTTAGACCG CATTGGACGT ATTTTAACGC CAGTCTTTGC 3720

AATTTTGATT GTTATCTTGG TCGTTCTGGG AGCTATCAAA TATGGTGGAA CAAGTCCTCA 3780

AGCTGCTTCA CTGCTTATCA AGCTTCTGCC TTTGGTACAG GTTTCCTAGA AGGTTACAAT 3840

ACCTTGGACG CCCTTGCCTC AGTGGCCTTT AGCGTAATCG CAGTTCAAAC CTTGAAACAA 3900

CTTGGATTTT CAAGTAAGAA AGAATACATT TCAACTATTT GGGTTGTTGG TATCGTTGTT 3960

GCCCTTGCCT TCAGCGCTCT TTACATCGGT TTAGGTTTTC TTGGAAATCA TTTCCCAGTA 4020

CCAGCTGAAG CGATGAAGGG TGGAACACCA GGTGTTTACA TCTTGTCACA AGCCACTCAA 4080

GAAATCTTTG GCTCAACAGC TCAACTCTTC CTTGCAGCTA TGGTTACCGT AACCTGCTTC 4140

ACAACGACTG TTGGTTTGAT TGTGTCAACA GCTGAGTTCT TTAATGAGCG CTTCCCACAA 4200

ATCAGCTACA AGGTTTATGC GACAGCCTTT ACCTTGATTG GATTTGCTAT TGCCAATTTG 4260

GGTCTTGATG CGATTATC 4278

(2) INFORMATION FOR SEQ ID NO: 154:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1953 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

ACCCGATCAA ATGACAAAAG CTAACTTTGG TGTCGTAGGT ATGGCCGTAA TGGGTCGTAA 60

CCTTGCCCTT AATATTGAAT CTCGTGGTTA CACAGTTGCT ATCTACAACC GTAGTAAAGA 120

AAAAACGGAA GATGTGATTG CTTGCCATCC TGAAAAGAAC TTTGTACCAA GCTATGACGT 180

TGAAAGTTTT GTAAACTCAA TCGAAAAACC TCGTCGTATC ATGCTGATGG TTCAAGCTGG 240

ACCTGGTACA GATGCTACTA TCCAAGCCCT TCTTCCACAC CTTGACAAGG GTGATATCTT 300

GATTGACGGA GGAAATACTT TCTACAAAGA TACCATCCGT CGTAATGAAG AATTGGCAAA 360

CTCTGGTATC AACTTTATCG GTACTGGGGT TTCTGGTGGT GAAAAAGGTG CCCTTGAAGG 420

TCCTTCTATC ATGCCTGGTG GACAAAAAGA AGCCTACGAA TTGGTTGCGG ATGTTCTTGA 480

AGAAATCTCA GCTAAAGCAC CAGAAGATGG CAAACCATGT GTGACTTACA TCGGTCCTGA 540

TGGAGCTGGT CACTATGTGA AAATGGTTCA CAATGGTATT GAGTACGGTG ATATGCAATT 600

GATCGCAGAA AGCTATGACT TGATGCAACA CTTGCTAGGC CTTTCTGCAG AAGATATGGC 660

TGAAATCTTT ACTGAGTGGA ACAAGGGTGA ATTAGACAGC TACTTGATTG AAATCACAGC 720

TGATATCTTG AGCCGTAAAG ACGATGAAGG CCAAGATGGA CCAATCGTAG ACTACATCCT 780

TGATGCTGCA GGTAACAAGG GAACTGGTAA ATGGACTAGC CAATCATCTC TTGACCTTGG 840

TGTACCATTG TCACTGATTA CTGAGTCAGT GTTTGCACGC TACATTTCAA CTTACAAAGA 900

AGAACGTGTA CATGCTAGCA AGGTGCTTCC AAAACCAGCT GCCTTCAACT TTGAAGGAGA 960

CAAGGCTGAA TTGATTGAAA AGATCCGTCA AGCCCTTTAC TTCTCAAAAA TCATTTCATA 1020

CGCACAAGGA TTTGCTCAAT TGCGTGTAGC CTCTAAAGAA AACAACTGGA ACTTGCCATT 1080

TGCAGATATC GCATCTATCT GGCGTGATGG CTGTATCATC CGTTCTCGTT TCTTGCAAAA 1140

GATTACAGAT GCTTACAACC GCGATGCAGA TCTTGCCAAC CTTCTTTTGG ACGAGTACTT 1200

CTTGGATGTT ACTGCTAAGT ACCAACAAGC AGTACGTGAT ATCGTAGCTC TTGCGGTTCA 1260

AGCAGGTGTG CCAGTGCCAA CTTTCTCAGC AGCTATTACT TACTTTGATA GCTACCGTTC 1320

AGCTGACCTT CCAGCTAACT TGATCCAAGC ACAACGTGAC TACTTTGGTG CTCACACTTA 1380

CCAACGTAAA GACAAAGAAG GAACCTTCCA CTACTCTTGG TATGACGAAA AATAAGTAGG 1440

TCAGCCATGG GGAAACGGAT TTTATTACTT GAGAAAGAAC GAAATCTAGC TCATTTTTTA 1500

AGTTTGGAAC TCCAGAAAGA GCAGTATCGG GTTGATCTGG TAGAGGAGGG GCAAAAAGCC 1560

CTCTCCATGG CTCTTCAGAC AGACTATGAT TTGATGTTAT TGAACGTTAA TCTGGGAGAT 1620

ATGATGGCTC AGGATTTTGC AGAAAAATTG AGCCGAACTA AACCTGCCTC AGTCATCATG 1680

ATTTTAGATC ATTGGGAAGA CTTGCAAGAA GAGCTGGAAG TTGTTCAGCG TTTTGCAGTT 1740

TCATACATCT ATAAGCCAGT CCTTATCGAA AATCTGGTAG CGCGTATTTC GGCGATCTTC 1800

CGAGGTCGGG ACTTCATTGA TCAACACTGC AGTCTGATGA AAGTTCCAAG GACCTACCGC 1860

AATCTTAGGA TAGATGTTGA ACATCACACG GTTTATCGTG GTGAAGAGAT GATTGCTCTG 1920

ACACGCCGTG AGTATGACCT TTTGGCGACA CGG 1953

(2) INFORMATION FOR SEQ ID NO: 155:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6474 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

CCGGCAGTAC ACGAGCTTGG GGAACAGCCA CTGGAACGAT GAGGTGTGAG CTCAAAATAT 60

1022

CCTCCAGTTA TGTTTTTCCT AATAGTATAC CGGAAGAGTG AAAGGATTTT ATAATGGAGC 120

GGTTACAAAG AACCTACTTT CTATTAAACA GTATACTATG AAAATGTGAA AATTTAACAT 180

TTTTTTGTAC AAATTTTATA AATTATTGCC TTTTTAATAT CAATAGTTAA TCTCTTATCC 240

AGATCCCCCT TGTGTAAACT TTATCTTTAT AAGCTTCAAG GCCCCTATCC CATCTATTTG 300

CAACAATTAG ATCACTTTGT TTTGTAAATA GTTCAAAATT CTTTTCAATA ATTACGTTAT 360

CTATACTAAC GTTTAAATTT GGTTCATATA CTAAAATTTT TATACCGACA ATCAATAGTT 420

CATTAATTAT ACTTAAAATA GCTGACTCTT TGTAATTATC TGAATTATAT TTCATCCCCA 480

ATTTATATAT TCCTACTATC TTTGGCTTTC GTTCCAATAT TTGTTTAACT ATGAACTGTT 540

TTCTATTTGT GTTTGAAATA TCAATCGCTT CTATCACTGG GGCATTTATT TCTATAAATT 600

CTTTTTTTAA TTGTTTAGTA TCTTTGGGAA GACAATATCC TCCAAATCCA AAAGAAGGAT 660

TATTATAAAA ATTTCCAATT CTTGGATCTA AACAAACACC TTTTATTACA ACTTCAGCAT 720

TTAAGCTTCT CCTCTCAGCA AAAGAATCTA GTTCATTAAA AAAGCAACAC GGAGAGCTAA 780

GAATGTGTTA GAAAAAAGCT TAATTGCTTC TGCTTCAGTA GGAGAAACTA ACATAACATT 840

TTTAATATTG GCAGTACTAT GAGTACTAAT CGAAAGGAAC AACTCTGCAA TTTTTCTTCC 900

TTCAACTGTC TCATCTCCAA CAACTATGCG ACTTGGATAT AAATTATCAT ATATAGAACA 960

ACCTTCTCTC AAAAATTCAG GGACAAAAAT GATATTTTTT GTATCAAACA GCCTTTTTAA 1020

TTTGTTTGAA AAGCCGATCG GAACTGTTGA CTTTAAAATA ATCTTTCCAT TAGGTTTTAC 1080

CCTCAGAATC TTCGATACCG TTTGTTCGAT TTCATATGTA TTAAAACTAC CAATTTTCTC 1140

ATCATAATCT GTCGGAAGCG CAATAATATA ATAATCAATA TTATTTTTAA TTTCAGAAAA 1200

TGTATCAAAA AAAGTAATAT TTAAGTTATT CTCGCAAAAA AACTTCATAA GCTCTTCATT 1260

TTTAGATGGA AGAATGCCCT TTTTTAAATT ATTTATTTTT ACAGAATCTA TATCATATGC 1320

AACAACTTTA TATTTAGATG CAAATAGTAA CGCGTAGGCC AGCCCAACAT GCCCCAAACC 1380

AATTACTGCT ATATTCATAA AACTACTTCC TTATTTCTTA ATCCAAAATC TAATAGAATA 1440

AGCTGCCCCA TTCCTTAAAT ACAACTCTTT AATATTGTTT AAAAGTTTTT CAACTGATTT 1500

CCAGATTATC AAAATCTGAG ATTTATAGCA CAATATTGAT GATATTCTAT CAATATAATT 1560

TTTTTCATCA AGTTCCTCTT GATACATTTT TAATTCTTTA GTTTTTCCCA TATAACTAAC 1620

CATACTACTA TCACTTACAT ATGGGAAGTC CTCATAATAT ATTACTTTAT AACGCATAAA 1680

TTCAAGCGCC CTTCCAATAC TATTCACAAA AACATGAGCA ACATGGTCAC CAAGTGAAAG 1740

CGGACAATAT ACGACACATT TGTCGTCTAA ATGCATTAAC AGCTCTTTTA TGATATCATT 1800

CTTTAATGTG TCCTCATTTT TTAATTCACT ATAGATATGA CGGTATAGAA AATTGCCATT 1860

TCTATCTTTC CTATAGAGAC ATTCATAGTA CGATAAGTGT CTAAAATCAC ATTGTAGACG 1920

TTCACAAGCT AACCTGTCTT CTTTCTTCCT TTCTTCAATC GGATATTTCC CAAGGTTACA 1980

CAACTTATGA AATTGCTTAG CAGAGGGCTG TAGCTGTTGG CTCAAAGGGT AACCAGAAAA 2040

TATAGTAATA ACAAGTACAA TTTCTCCTTC TGAAGTTAAT TTTGAAATAT AATCACCACA 2100

GGAAAAAATT GCGTCATCTA AATGTGGAGA TAAAAAGATA TACTTAGTAT TGTTACTCAT 2160

AACCATTCCC TCTACAATTT ATCTAAAAAC TCACTAAGTG TCTGATTAAA TTCCACATCA 2220

TCAAAAAAAT TCACCTTATT CTTAATAATG AATATTTCGT TAAATAAACA TATATATAAA 2280

TATTTCAATA TCCTTTCAAT ATCATCCTCT AAATTCTCCT CAATATTTTG TATCAGCCCA 2340

TTTACAATCT TATTAAAAAA GATAAGCTCT TTATCTCTAA AATTAAATAT TTTCATACAA 2400

CTGTTGTATC GAAAAATATA TAAAATAATT TTTACTAATG TTTGAATATT TAAACAACTA 2460

AATAAATGAG TTGTACCCGG GACACTATTT ATGTTATCAA GAACACTATC TTGAAACCTC 2520

AACTCACAGT TCTTTTTGTG AAATTCTTTT TTATCGTTTA GATCTGATAT TTTTTTAGAC 2580

ATTTCAACAA TCTCAGACAT TTTATATGGA TATCTAGGAT GAATGCCAAA ACTATGCAAA 2640

ATGAACTGCA CCCCAAAAGT TAGACAGAAT AAATCTAACT TTTGGGGTGC AGTTCATAAG 2700

ATTGGGATAT TTTTTTTTAG CTAGAACTAG TAGAAATATA TAGTCAAATA ACAGATACCT 2760

TAAGGGTTTC TCATCTACAT AAAAAAATGA TACTTTTTTC TCTTCAGTAA TTACCTCATA 2820

AGCTTCACAA TAGAATCTCA TGTTTCCCTC CCCTATATTC TTAAATAAAA TCCTTTGGAA 2880

ATTGATATAT CTTAGTAAAA TATTGTTTAA GTTCCGGATG CGGAGCATGG GTAACAATAA 2940

TGACAGTCAA ATCCTCTCTA TCTAATATCT TACGTTCAAT CGCTAACGAA GTTCTCCTAT 3000

CGATAGCAGA AGTTCCCTCG TCAATTAATA CTATTTTCTT ATTTCTAATT AGCCCTCTAG 3060

CTAAAGTAAT TTTTTGTTTC TGCCCTCCTG ACAGTAATCT CCCATCATCA CCAACATAAT 3120

AATCTAAAAT GTTATTAGGA AAATCTTTTA CACTCAAACC AACTTGCTCT AAAGACTGTA 3180

GTATTTCTTC ATCAGTATAA TTTTCTTCCA ATAAAATATT ATCTCTAATC GTACCTTCAA 3240

ACAAATAAGC TTTTTGATCT ACATATAGAA CATTCGAAAC CATATTTAAA TAGGAGGTTT 3300

TTTTTATATC ATCCCCGCAG AATCGCAATT CTCCACTATA ATCTCTCAAA AAGCCATTCA 3360

ATAATTTTAA TAATGTAGAT TTCCCGCTTC CACTTTCACC TAAAATTAAA TACTTTTCAT 3420

TACGTTGAAA ACAAAAATTT AAGTTTTTTA ATATTTCTTT ATCTCCATAC TTATAGCAAA 3480

TATTTTTTGC TTCATATAAC GGAAAATCTC TATTCACCTC ATTTGGTTCG ATATCATTCA 3540

TTTTATTTGA CTCAATTGGA TTAATTGAAT ACAATTTTAA AAAAATAGGC TTCGTACCAA 3600

TAATAGAGGA TAATTGACCT CCTAATTCAC CTAGCGCTGT AAAAATAACA CCTGTTAGTG 3660

CTCCTATTGC TTCAATAGTA CCAATTTTCA CTATTCCTTT TATTGCAAGA TAGCCTGTTA 3720

AAAAAACGAG AGATATCTGA AAAAAAATAT TGAGAAAGAA GCTAATAGCG CCTGCTAACG 3780

TTTCTACAGT TGTCTTTCTT TGTATAACCA TCTTTAATAA AATTCCTGCT TCTTTAATTT 3840

TCTTAGGCAA TACATATAAA AGATTCAAGG ACGCTAACAC ATCAAATCCA TTCAATATAG 3900

TCTCACTAGA TTTTAAAAAA GCTTCATTTT GGTTAGTTAA ATTTAGACTA ACTTCTCGCA 3960

TTTTCGATGC AAAGATTTTT GGTACAAGTA GCATAATCAT TAATGAAAAC AAGGTGGCTA 4020

CAGTCAATGA CCAATGATAG TGATTAAGAG TCACAACTGC AAATATAGTA CCAGAAATTC 4080

CTTTTATTAC TAAAAAAAGT TGTTTAAACG CCTGATCATT TAAAGTCTGA ACATCATTAT 4140

TTAGCCACGA AAGATATGTT CCTGATGATT TACTATGAAA TTCTTGATAG GTAGAGTTAG 4200

AGATGTCTGT GGCAACTCTA TTTCGAATCT CTAGATTAAA CTCTTGGATC ACTTCAACCT 4260

GATAATTTTT CACTACCCAG TCAAGGAATA TTATCCCACA CCAGACAATC ATTTGGTAGA 4320

TTGACAATTT CAAAAACCGC TCTAAATTCA TCGCAATTAA TTCATTCAAC ACCAGAGCAT 4380

TAATAGTTGC TGCATAAATT AGCAATAATT GACCAGCAAC AATAAATATC GTTAATAAAC 4440

TAAATTTTTT TATATTTGAT TTTATAATAG TATACACAAT AGTTTCTCAC TTTCTAAATT 4500

TTAATTGAAC ATAGTTTTCA TATATACAAT AGAAAAAACC AAAATGATAT AATAACATAT 4560

ATTTCAAAAA AGAAATTCGT TAAAAATTTT TTCTTCTCTT GCCTTCTTGA TTACTTTTAA 4620

AGCCTTGCAT TTGTCTCCTA TTAATAGTAA CCGCTTTATG TTTAAAGAAT AATATTTCTT 4680

TGTAACCAAT ATTCTCTCGT TGAAACTCAA TAAATTAAAA TATTTCCTAC AGTAATTATA 4740

ATATTCTTCA TCTGCATTAA TTGTTTTTTG TGTCACTCCA GTGATACCGT TTTCTTTACT 4800

GTGAGCGTAG TAATTCACCA AGAATTCTCG CACTATATCA ATTTGGTATC CTTGAACAAG 4860

TAGTTTTAAT AAAACAACAC CGTCCTGATG TGAATCTATT TTCTCAAAAC CATTAATTAA 4920

TTCTAGCACC TCTTTTTTAC ACAACCAAAA TGACGTACCT GCTATATTGT GAACCATTTG 4980

AACAAACAAG GGATTTCCAA CAAAATCGGT CTTCTCCTCT TCTCGTGTAC CATTTGGATA 5040

AATTATTATT CCATAACTAC AAACTAAAGC TAAATTCTTC ATTCTACTCT TTTTAAAACA 5100

AGCCATCAAC TTTAAAATTC GATCTGGCAT ATATTCATCA TCATCGTCTA AAAATGATAT 5160

ATACTTACCT CTAGAATTTT TGATACCTAT GTTTCTGGCA TTAGTTGCAC CTAAATCTTC 5220

ATTACTTAAA ATTAACTTAA TTCTATGATT GGTATAGCCA AATTGATGGA TAATTTTATT 5280

TCTTAAATTT ACATTACTAT AATTATCATC AATAATTATA ACTTCGATAT TTTTATAACT 5340

TTGATGTAAA CAACTTTTCA CAGCTCTAAT CAGAGATTCA TACCTATTAT GTGTTGGTAT 5400

TATAATACTT ACTAATTCTT GATCTATATT CCTATCCATG ACTACTCTTC TCTAATAATT 5460

CATCATATAC TCTCATGGTT TCTACAAACA TTTTTTGCAC AGAAAAATGT TTTCTTATTT 5520

TTGATTTACT ATTCTCACCT ATATATTTCA AATACTCAGA ATCATTGAGT AAAAAATTAG 5580

CACAAGCACA CACTCCCTCA ACATCTTCCT TCTCAAATAA AAATCCATCA ACCCTATGTT 5640

CAATAATTTC ACTTAACCCG CCAACATTAC TAGCTAAAAC CGGAGTTCCT TGTGACATTG 5700

ACTCTAAAAC ACACATAGGT ATTCCTTCTG TATCAGAAGG AATATACAAT AAATCCGATA 5760

TTTGGTAAAC TATAGTAGCT GGATAGATTT CACCAAGTAA CCTGAAATTA TCTCTACATT 5820

TCAAATGGCA AATTTTTTCT TTCAAAGCAG CCCACATACT ACCATTTCCA GCCATAATAA 5880

AAATCACATC TTCTCTGACT AAAAATAATT TTTCTGCAAA TTCAAGGAAT CTATCCGGCC 5940

TTTTTTCTGG ATCCAACCTT CCAACATAAC AAATGATTTT TTGTTATTTG GAATACAAAA 6000

TTCTTTTTTA AAGTCTTGAA CACCTACTAC ATCTAAATCG CTATTTGATA CATTAATTCC 6060

GTTATTTATT GCAACTATCT TCTTATTTTT TATTATACTC TCCAATCTTT TTTTTCATAG 6120

TTTCAGATAC ACAAATAAAA GCATCTCCCA TAGAATATGT CCAAAAATCA AAATAAGTCA 6180

AGAATTTCTT TTTTAAGTTA TATTCAACCC ATCCATGGCA TGTTATCACT GTCTTAACCT 6240

TTCCAAATCC ATTCTTGTCA AGTTTTTTTA ACATATATAA AAAATAATTA GTTGAGTAGC 6300

CATGACAGTG TATAAGTTGG ATTTTTAATA ATTTTAAAAT ATTTTTAACG TGTAAGGCAG 6360

TTTCAAAATT ATTTGAACAT TGAGTACAAT CAACATAGGC AATATCTAAA TTTTTATAAT 6420

CATCAATAAC CTTTGAATCT CTAGATACAA TTATCAAAAT AGGGAATAGA GACA 6474

(2) INFORMATION FOR SEQ ID NO: 156:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4792 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

TATTTAACGA TTTTTTTCAT GTCATTTCCT CCAAAATAGA ATACCTTATA ATCTTAACAG 60

AAAAAGAGCA TTTACGCCAT TATATGATAT CTATCTCTGT GATAAGTTTT TTTTATGGGT 120

AATTTAAAAG ACCAAACGCA AGATGGCAAT CAAGACCACT CCAAAGAGAA CTGTTCCGAC 180

TAGATTGCGG TAGCGAAAGG CTACCCAAGC TGTTGGAAAG ACGGCTAAGA AGTCCAGTCA 240

TTTGATTTGA GGAAGACTGC CAACCTTACC TGTCACTACG CTTGAAAGAA TCAGGGCAAA 300

GATAATGGAA ACAGGCAAAA ACTTCAAAAA ACGCTCAACA ATCGCAGGCA GGCCCTTATA 360

CTTGACCAAG ATGAAGGGAA TCATACGGGG AATCCAAGTC ACCAAGCCAG AGAAAATAAC 420

TGCTAATAAA AGATACTTAC TGACCATCTA AAACCACCCC CATGCTACAA CCAAGTAGCG 480

TCGCAAACAG AACAGCTAGT GACTGAGACA TCACTGTCAA GAGCAAAAAG AAGGACACCG 540

CAACAACTGC TAGGATAATG AGCAGATTGC GGACAGGAAT CCGTCTTTGC ATAATCTGAA 600

ATTGCGAAGC AAAATACCAA TAAACATCCC AACCAGGGCA AAATCCAAGC CAAAGATTTC 660

TGGATTTGGT AGCAGGCCAC CCAGAGCCGT TCCGACTACT GTCCCCACAA ACCAAGCCAC 720

ATAGCTGTTA AGATTGTTTC CGTGCATCCA CATAGGATTT ACCTTGTCTG TATGGGCCAA 780

TTCACCCATC AAAACGCCAT AGGTCTCATC TGTCAAGATA CTAGACATAC CGATATTGTA 840

CCAAAGACTG GTATGACGGA AATAAGTCGA TGCGTGTAAA CTCAACAAAA AGAGACGCAA 900

GTTGATTAGA AAAACCGTCA TAGCAATAGC TGCCACAGGA GCTTGAACCA CAATCAGTGC 960

CAACATGGCA AACTGGGCAC TCCCAGCATA AACAAAGAGA CTCATCAAGC CCATCTCAAC 1020

AGGTGTCACA TAGGGCGCAC CGATAATTCC ACAGGCCAGG CCGATACTGA CATAGCCAAG 1080

AGCCGTTGGC ATGGCTGCCT GCGCCCCCTC CTAAAATCCT TTTTCTTTCA TCTTTCTCCT 1140

CATATTGTCT TAATAATACT CAATGAAAAT CAAAGAGCAA ACTAGGAAAC TAGCCGCAGG 1200

TTGCTCAAAA CACTGTTTTG AGGTTGCAGA TAGAACTGAT GAAGTCAGCT CAAAACACTG 1260

TTTTGAGGTT GTGGATAGAA CTGACGAAGT CAGCTCAAAA CACCGTTTTG AGGTTGTGGA 1320

TAGAACTGAC GAAGTCAGTA ACCATACCTA CGGCAAAGTG AAGCTGACGT GGTTTGAAGA 1380

GAGTTTCGAA GAGTACAAGT AGGCTGAAAA GAATCCAACC ACAGCATGGA CTATTATATA 1440

GCAGATTGAA ATAAGATGAG AACAAATCGA TTGGGAAAGT AAAATTAATT TCTATAAATG 1500

TTTTAGCAAT TGTTTCGTAC TATTTTAGAT TCAGTCTATT ATAACACATT CAGAAAAGAG 1560

AAAAAAGTCT GTTGATTTTG ACCATCATAA AAAGACTGGC AATCCAGTCT CAAACATATA 1620

TTATAGAAAT TCTCCACTAA ATACTTTCAC GAATATTCAG AAGCATAACA AAGGCAACTA 1680

GAAGAAATAG CAATAAAACA AAGCTAACTG CCAGAGTTCC AAAGCTAGTA GCAATGGTTA 1740

CCAAAGCTAT TGTAAATAAG CTAGGTAAAA CAACCGTAAT GGCACCGATA GAGGATTGAA 1800

CTGCTCCCAT TGACTCCTCA GGTATTTGTT TAAAAACGAG TTCTTGCAAT CTAGGAGAGA 1860

GAACACCTGC GAAAAAGGCA TCCAAGGTAC TAAAGATGAG AATCCAGTCA AAACGAACTG 1920

TGGCAAATCC TACTAGAAGA AGCAACTGGA TGACAAGTGA GGCATAGAGA GCTGTTTTTA 1980

TGGAAATGGT ATGTTGCAGA TAGCCACTTA CAAGGCTTCC GACAATCAGG GCTGATAATT 2040

CTAGTGTGGC TAACAAGGCA AGAGATTGAC CAGTTTGTAA ATTCAAAAAG GGCTGGTTCC 2100

TTAAAAATAG AGTGGAAATA GGAACCGTAA CATTTATCAC TGCTTGACTA GTAGAGATAA 2160

TAAACAAAAC CAAGAGCACC TTATTCATAT TCCATATCAA TTTCGATGAT TGGAGCAAAT 2220

GCTGGCAAAA GGATTTTACA GAGAGTCCTT CTTGATAGCT AATCGTTTTT TCTACTTTCA 2280

AGAGGTCAGT TTTTATGAAG AGGATACCTA AAAATGCGAT TAAAAAGGTA AGAGCGTTCA 2340

GTAAGGAAAT AAACTGGATG GATAGAATGC CTAGTAAGAC TCCTCCTAGG ATATTACTGA 2400

TTGTTTTCAC TAAACTAACA GTTGACTGTT TAAAGCCAAT AGCTTCTGCC AGATGGTCTT 2460

GCCCAATAAT TCTAATGAAA ATCGGAGTGA GCATGGCGCC TGAAAAATAA CTCAATGTGT 2520

CAGACAAGAG GTTAATCAGA CAAATAAATG CTACTAGCAA CAAGGAGAAA GACTGCCCTG 2580

AAAGTGATAA AGACACTATA GAGTAAAGCA AAAATTTTGC AAAACTAATG ACTGTGTATT 2640

TCAAGACACG ATGATGTTGA AAATCCGCCA AAACTCCCAG AAAGATTTGT AGAACTTGGG 2700

GCAGGGTTTC TGAAATCGTG ATGAGTAAAA TCGCCAAAGG GGCAAAAGAT GCATCTGCCA 2760

CATAATTCAG GAAGGCCAGA TAAAAAATCG TATCCCCAAG CGTTGAAATC CACTGGTTGA 2820

TAGTTAATTG CCTAAAATCT CTATTTTGAA GAAATACTTT CATCACAACT CCTTCTTAAG 2880

TTCAAATGGG AATCTTTCCC CAAGGATAGA CCGCGATACT ACTAACAACC AAAATTACAG 2940

TAACATCAAA AGCTGACCAA TGCCATTGTA GACTATATGC AGTCCAATAG GCCAATAAAT 3000

TGACTTTGTC ATTCTAAATA AGACTGCAAA TATAAGACCT CCACCCATAT AGAAGACAAA 3060

GTCTGTCAAG ACCCAACCGT GATTACTAAT GTGCGAGACC CCAAATAAAA CAGCGGAACC 3120

AAGTACATCT AGCCCCCATT TCTTTCCTTT TTCCAGAGCA GTCATCACTA ATCCACGATA 3180

AATCATGTCT TCAAAAATGG GACCTGCAAT CACAGGATAA AAAAAATACA TCAAAAATGC 3240

TGTAGCCCCT GTAAAAGTCG GAGCAGCATG TTGATAAGAA ATTTCATTTC GAGTAGGTGG 3300

GAAAAGAAAA AAGGTAACGA AATTCCAAAC AACAAAAGCA AGCAGAGCTA GGAAGGAATA 3360

GAAAAGATAG GATCCTTTAA ACTTTCTACT ATTGATTTTC TGCCATTTCC CCGACCAAAT 3420

CATAGCAATA AGAGCAAATA AAACCACAAG AAAATTCAAC ATCATATCCG ACAGATAATA 3480

GGCAAAGTCA GATAGCCCAG TAACAAGGTC GCTGCGTAAA ACTAGAACAC TGAACTTCTG 3540

GTCAGCAATA ACTAGTAGAA AAACTATAAT AAAGTAGCGG TGTGAGATTA TCTTTTTCAT 3600

ATATCACCTT TCTAATATCC AAATACCAAT AAAGTAACAA TGAGTAAGAA ACTATTCCAT 3660

GAAGCATGCA GAGCTATAGC CCAATAGATG GATCGGGTGT AGCGAAACAT CATACAAAAT 3720

ATCAAGCCCA TTCCAAAATA CTTTATGAAA TCTGTCGTTA TCCAACCATA CTGCAAAACA 3780

TGCATAGCGC CAAATATGGC AGCGGAAACA AGAACATCAA GATAGTATCT CTTAACTTTA 3840

GATAAACTTG TCATCAAAAG ACCACGACAA ACAACCTCTT CTGATACAGG TGCGATAATA 3900

CTAGTATAAA GTATTCGCGT AACAAAATAG CTAATTCCTG TTAAATTGGT GGCTACTTCT 3960

ACGACTGTAC TTCCATTCTG GGTACGAGGA AAGATATAGG TTGTTAGATT TGCCCACACG 4020

AACAATAAGA AAAAAGAAAG AAGGAAAACA CCCAGGTAAG ACCAACGAAA CTGGAAACGA 4080

CCACACTCTT TCCAATGTTC ACTTTTGACA AAAGCAATTG TAGCTATAGT TCCCAGAATA 4140

AGTACCAATA AAACTTGGAA CACATAGTAC ATATTATCAG ACAAAGCAAC CATAAAATCT 4200

AAGTCTGATG TGACATTAAA AATGAGGTAA TAAGTCAAAA TCAACAAGCC AGTTGCTAGG 4260

TGAAATTTCA CTTCTTTCAT TTTCTTCATC CTATTATCTC CTATAAGAGC CTATCTTCTA 4320

CGGCGGCCAA ACAATCCATC TGCTAAATCT ATAGTCCAAT CAAAAGCTCC ACGATTAGGA 4380

CTCATCCCTT GATTGCCCCA ACCAGGGTAA ATTCCTGGGA CGCCCCAACC AGATATACCA 4440

CTTCTTCCAC CACCTCCCAT AGAATTTACG AGGTTGCCTC CTCTAACATC TTGCAACTCA 4500

GCTTCTGTCA ATTCCATTGT TTCTGCAAAT TGTAAATTTA ACATCTTTTA CACTCCTTCA 4560

ATTATCTTCA TTTGTAAACC ACTTCTGCGA CCTAGGATTT GCTTCAAGTG CTTTACAAGT 4620

ACAGTATAAC ACGAACATTG GCTTATTTTA GAAAATCGCA TATTTGATAT TTTTTCTTAT 4680

AGAAATTTCA GATTTGCGAT TTTGGTGAAT TTGATTACTT CTCTGGTATA ATAAAGTTAC 4740

TACTAATGAG GAGTGGAGAA ATATGAAGAA ACAAATTTTA ACATTATTGA AA 4792

(2) INFORMATION FOR SEQ ID NO: 157:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2156 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

CCGTTCTCGG CGACGGCCAT CTGATGAAGC TATTTATGAG GGAAACTGGC AAGCTGGAGA 60

GTCAGAGTAT CTAGTCTTTC ACCGATTGCT GTGGCAGCAG ATGTGCAGGG AAAAGGAGTT 120

GCTCAAACCT TCTTAGAGGG CTTGATTGAA GGTTTTGATT ATCTTGATTT TCGCTCAGAT 180

ACGCATGCTG AAAACAAGGT TATGCAACAT ATTTTTGAAA AACTTGGTTT TAAACAAGTC 240

GGTAAGATGC CAGTAGATGG CGAACGCTTG GCCTATCAAG AATTAAAGAA ATAATGCAAA 300

AGAAGTATGT AAAAATCCTC TACTCCTCAC CAATTGGTAT TCTATCACTT GTAGCTGATG 360

ACCATTATTT GTATGGAATT TGGGTTCAGG AGCAGAAGCA TTTTGAGAGG GGACTAGGAG 420

ATGAAACGAT AGAAGAAGTT GTTAGTCATC CTATTTTAGA CCCAGTTATT GCTTGCTTAG 480

ATGATTACTT TAAAGGCAAG CCTCAGGATT TATCCAACTT GCTCTTGGCG CCAATCGGAA 540

CGAATTTTGA AAAGAGAGTT TGGGACTATT TACAGGGCAT TCCTTATGGT CAGACAGTGA 600

CCTATGGACA AATTGCTCAA GACCTGCAAG TGGCTTCTGC TCAAGCAATT GGTGGAGCAG 660

TGGGACGCAA TCCTTGGTCT ATCCTAGTAC CTTGTCATCG TGTGTTGGGA GCAGGCAAGC 720

GTCTGACAGG TTATGCTGCA GGAGTGGAAA AGAAAGCTTG GCTCTTGGAG CATGAAGGAG 780

TAGATTTTAA AGATAGAAGC AATAGAAGGA GAAGCACATG TTAGAATTTA TCGAATACCC 840

CAAATGTTCA ACTTGTAAAA AAGCAAAACA AGAATTAAAT CAATTAGGTG TGGACTATAA 900

AGCCGTCCAT ATCGTGGAAG AAACACCTAG CCAAGAAGTC ATTTTGAATT GGCTAGAAAC 960

CTCAGGATTT GAATTGAAGC AATTTTTCAA CACCAGTGGT ATCAAATACC GTGAATTAGG 1020

GCTAAAAGAT AAGGTAGGAA GTTTGTCAAA CCAAGAAGCG GCTGAGTTGC TAGCAAGTGA 1080

CGGTATGTTG TTAAAACGGC CCATTTTAGT AGAAAATGGA ACTGTTAAGC AAATCGGTTA 1140

TCGAAAATCT TATGAGGAAC TGGGACTGAA ATAGTTTTTA TCTATCTCTT TGATAGATAA 1200

AATATATAAC TTCCCTGTTT CAAAGTATGA TAAACTAGTA GGTAGACAAA GTCTGTATCT 1260

GACCGTAGCA AATAATTTCA TTGACGGCAG AAGCATGGTA GCATGAATCA TTATCAGAAG 1320

AGGATGTTTT TATGAATGTT ACAACGATTT TAGCATCAGA TTGGTACCAA AACTTGATGC 1380

AATTGATTCC GGATGGCAAG CTGTTTAGCC TACGTTCGGT CTTTGATGGA ATCCCTAGAA 1440

TTGTCCAACA ACTTCCAACA ACAATTATGT TGACAATTGG TGGTGCCCTT TTTGGCTTGG 1500

TTTTGGCGCT TCTTTTTGCC ATTGTGAAGA TCAATCGTGT CAAGATTTTA TATCCCTTGC 1560

AGGCCTTCTT TGTTAGTTTC TTAAAAGGGA CACcGATTTT GGTGCAACTC ATGTTGACCT 1620

ACTACGGAAT CCCTTTGGCT TTGAAAGCCC TCAATCAGCA ATGGGGAACT GGTCTCAATA 1680

TCAATGCGAT TCCAGCTGCA GCTTTTGCGA TTGTCGCCTT TGCCTTTAAT GAGGCAGCTT 1740

ATGCTAGTGA AACCATTCGT GCAGCCATTC TCTCAGTTAA TCCTGGTGAG ATTGAGGCGG 1800

CACGCAGTCT GGGTATGACC CGAGCGCAAG TTTATCGACG AGTGATTATT CCTAATGCAG 1860

CGGTGGTAGC TACTCCAACC TTGATTAATT CCCTCATCGG TTTGACCAAG GGAACATCTC 1920

TAGCTTTTAG TGCGGGTGTT GTGGAAGTCT TTGCCCAAGC TCAGATTCTA GGTGGAGCTG 1980

ATTATCGCTA TTTTGAACGC TTCATCTCCG TTGCCCTTGT TTATTGGGTA GTCAATATCG 2040

GAATTGAAAG CCTCGGTCGT TTCATCGAGA GAAAAATGGC TATTTCTGCA CCTGATACAG 2100

TGCAACAGAT GTGAAAGGAG ACCTTCGTTA ATGATTAAGA TTTCGAATTT AAGCAA 2156

(2) INFORMATION FOR SEQ ID NO: 158:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3140 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

GTATCTCTAC ACATGTCTTC AATCGATTTT GTTGTCCTCC AATTTAATTC CTTATATGCT 60

TTGTCTGCAT TTGCATAACA AGTTGCAACG TCTCCTGAAC GTCTTGGAAC TATTTTATAA 120

GGAATAGGGA TCTTATTAAC ACTTTCAAAT GTATTTACAA GTTGTAATAC ACTAGTGCCT 180

TCTCCCGAGC CTAGGTTATA GATATAAACA TCTGTTTTTT CAGATACTTT TTCTAAAGCT 240

TTTATATGTC CTATTGCTAA ATCTACTACA TGGATATAAT CACGCACACC AGTACCATCA 300

AGCGTATCAT AATCATTTCC GAACACACTT AGCTCTGATA GCTTACCTAC CGCTACTTGT 360

GCAATATAAG GCATCAAGTT GTTAGGAATT CCTGAGGGAT CTTCCCCAAT CAAACCAGAC 420

TCATGAGCAC CAATTGGATT GAAATAACGA AGCAACGCAA TACTCCATTC TGAATCTGCC 480

ACATGAACAT CTTTTAAAAT TTGCTCAAGC ATCACTTTCG TATACCCATA AGGATTTGTC 540

GCACTTGTTT GCATCGTCTC AATTAGAGGT GACTGATTGT TAATTCCATA TACAGTCGCA 600

CTTGAAGAAA AGACAATCTT TTTAACATTA AATTCTGACA TCACTTCAAC AAGTGCCAAT 660

GTACTCATAA TATTATTTTT GTAGTACATC ACAGGCTTTT GCACGGATTC TCCGACAGCT 720

TTATAACCTG CAAAATGAAT TGCAGCATCA ATCGATTCTT GTTCAAATAC CTTTCTCAAT 780

GCTTGTTTAT CACAAACATC TAATTCGTAA AACACGGGAC GTATTCCTGT AATTGCTTCA 840

ATACGGTCTA GCACCAAGAT GCTAGAGTTC GAAAGGTTGT CGACAATGAT AACTTCCTTT 900

CCTAAATTTA GTAATTCTAC TACGGTATGG CTACCAATAT AACCAGCTCC GCCTGTTACC 960

AATATTGCCA TCTGGGTTTC CTCCTAATTA ATTCCAACCG ACTTAACAAA TCTCATAAAC 1020

GCTTCATGCC CAGACGGTGT ATTCTTATAA ACTCCTGCAT CTTCCAGAAC TCTCGCAAAC 1080

ACTTGTCCTG CTTCGTGTTG AACTACGCTA TTAACCTCTT CTTTATTAAT GCGAGGATAT 1140

TTTTCTTTCA ATTGGTCGGC CCATTCTAAA TGATAATCCG CAATTGCATT ATCCTCTCCT 1200

AAAAGATATT TTCCAACTTC TTCTAACTCT GGTTTCAAAC GAGGTGGTAA TATCGCAAGT 1260

CCCATCACTT CGATTAACCC GATATTTTCC TTTTTAATAT GTTGTACATC TTGATGAGGA 1320

TGGAAAACAC CATCTGGGTA TTGTTCAGTA GTATGATTAT CTCTTAGAAC AATATCTAAT 1380

TCGTATCTCC CGTCCACTTT ACGAGCAATA GGAGTCACCG TATGGTGTGG GACATCTTCA 1440

GTCATAGCAA TGATGTCTAC TTCTAAATCT GAATATTCTC TCCACTTATT TAGAATTTTA 1500

GTAGCTAAAT CTAACAAGCG ATTTTTATTT TCACTTTGTA ACCTAATTAC TGACATTGGC 1560

CATTTTACAA TACCAGCATT AACATCCTCA AAGTCTTTAA AACAAAATTC ACTCTCAAAT 1620

TTTGCTTTTT CCATTGGGAA AATATGTTTC CCTCCCTGGT AGTGGTTATG ACTAAGAATG 1680

GAGCCTCCTG AGATAGGAAG ATCAGAATTT GAACCAGCAA AATATCCTGG CAAAATATCA 1740

ACAATCTCCA ATAATTGTTC AAATGTTTTA GAGGTAATAG CCATTGGTAC ATGTTGACTA 1800

TTCAAAAATA TCGCATGCTC ATTAAAGTAT GAGTAGGGAG AATACTGGAA TCCCCATACT 1860

TCGTCACCAA GTTTCAACCG AATAATTCTA TGATTCGAAC GTGCTGGATA ATTTATTCGC 1920

CCCTGATATC CTTCATTTTC CATACATAGT AAACATTTGG GATAATTAGT TGCTTTTACT 1980

AATTTTTCAG CAGCAATTGT TTTTGGATCT TTTTCGGGTT TTGACAAATT TATCGTAATC 2040

TCTAGCTCTC CGTATTTAGT TGATGCTCGA AACTCAATAT TCTTAGCAAT AGCAGAAGTT 2100

TTAATATAAT CACTATCTTT ACTTAACTTA TAAAACTCTT CAACTGCTTC TTGAGGTGAT 2160

ATATCATATG AACTCCAAAA AATATCATTT AATCGACTAG GTAAAGGAAC TATGAAATTC 2220

ATTAACTCTG CTCCTAAACA TTCCTTTTCC TCGATTAAAT CTTTAATTTT ACCGTTTTTT 2280

AAGGCGATTT CCACTAAGTA ATCTTTTATT TGTTTCAGGT CATTTTCATC GGAAATGCGA 2340

TCAATTCCCT CCTCACCTAT TAACGCTAGT ACTCTATTTT TCACATATAT TTTGTCAATT 2400

TCATTATACA TTCCGTATTC AATTACTCTA TCAACAAAAT TATCAATAAT TGTTTTCATA 2460

TATTTTTCTT TCTAATTTAT GTTCCCATAT TTTCTATACA TTATCCATTT ATAAATTGCT 2520

TGCGTAGTAT GAGCAATTTT ATCAAGGTGA TGAATAATAT CTAAAGCACT AATTACTTCA 2580

GAAACGTTCC CATCATCTTC AAATATGTAA TTCATTATTT TCTTTTCCAT ATTTATACTA 2640

AGCTCTTCTA TCTCATTCTG TTTTTGTATA ACAACCATAT CTAAACATCC AGATTGTTCC 2700

TCTCTATAAC AAGATATAGC CCTATTCATA TGCAGTCCGA TAACTTCATG AAGTATTTTT 2760

ATTTTTGAAA TAATTTTCTT CAAAATTTCA TTATTTTGAA GAATCTGTAG ATTTTTTAAA 2820

ATTTCAACAA TTCTATCCCC AATACGTTCA ATGTCAGTTG ATATTTTTAT TACACTAATA 2880

ATTCTTCTTA AGTCATATGA AACAGGATGT TGTAAACAAA TTAACTCATA TCCTTTTTTA 2940

TCAATATTTA GAACTGACTC ATTTATGATT AAATCTTCTT TAATCAATTC TACTCGTTCT 3000

TCATTTGATA AATATTCAAA TAACTTCTCA TATTTATCAA GCACAGATAC CCAAATGGTC 3060

TCTAAATTAT TTGATAATTC TATAATTTCA TTTTCTAAAT ATAACCTTAA CATTTAGGTA 3120

CCTCTTCTTA ACAAAGTTCG 3140

(2) INFORMATION FOR SEQ ID NO: 159:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9048 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

CCGGATGATT TCCTGGTCAG ATAGGGGGAA AGTGACTTCC TCAGCAATCG CGCGTAGAGT 60

AGGATTCCCT TCACGGATAA TATCGTTCAT ATCAATTAAG TGAGCAGCTT TTGTAATACG 120

TTCTATTGCA GACATTTTCT CTCCTTATAT TATGTTTAGT GCAGTTAGCT ACTGCCAAAG 180

CCCAAGTGGT ATACTTGGAA TAAGCCACTG TGGATTAGTT CATTTTCTTT CATTACCTCT 240

ACATGATATC ACAAAATGAC AAGAATTGAA AGCATTATGG CATTTAGGAT TTATAGAAAA 300

TAGATAGGAA GTTCAATTCA ATTGTGAAAG AAATACTTAT CTGTGATATA ATAAAAAGAA 360

AAGGCTTGCA TAAGAAAGTA GGGAGAACGA AGATACAAAG AAGACAAAAT CGAAATCAGG 420

GTGGTTTAGC TTTTCGTTTT ATGAAGGGCT TGGTAAACTT TTTAGGAGTT ATCGCAAGTG 480

GAGCAATAAG GGATTTGTGG CGATACTCTT GCTAGCAGTT GGTTTATCAA TGGGCTTGGT 540

CTTGTTGTTT GAAAGCTTCC AAGGAATCCC TTGACTAGTC AAAAACGAGA TACTATTTCT 600

CAAGAGGGGA CTAAGCAAAA GTCTCAGGAG TAGGAAGAGG AAAAAACTGC CAGAATTATG 660

GCCCACGGGG ATTTGCTCTA CCACGATGGA CTTTTCTTTT CAGCTAAAAA AGAAGACGGT 720

ACCTATGACT TTCATGAAAA TTTTGAGTAT GTGACTCCTT GGCTCAAGCA AGGGGACTAA 780

GCAGCAGATT TAGCTATTGG TGATTTTGAA GGAACCATTA ATAAGGATCA TTATTTAGCG 840

GGTTATCTTC TCTTTAATGC TCCTGTTGAA GTTATGGATG CTATTAAGGA GGCAGGTTAT 900

CATGTGCTGG ATTTAGCTCA TAATCATATT TTGGATTCGC AAATTGAGGG AGTTATTTCA 960

ACGGCCGATA TTATTGAGAA AGCTGGAATC ACTCCAATCG GAGTTTATAC GCACGAACCA 1020

CGTGATCAGG CTCCGCTGGT CATTAAGGAA GTGAATGGTA TCAAGGTTGC ATTGTTAGCC 1080

TATTCCTATG GTTTCAATGG AATTGAGCAG TATATTTCTC AGGAAGACTA TAATCGTTAT 1140

CTTTCAGATT TAAACGAAGA TAAGATGAAG GTTGAAATTG AACGGGCAGA GAAGGAAGCA 1200

GATATCACCA TTATCATGCT TCAGATGGGT GTTGAGTATC GATTGGAACC AACTGAAGAA 1260

CAAAAAGCTC TTTATCACAA GATGATCGAT TTGGGAGCGG ATATTATCTT TGGAGGGCAT 1320

CCTCACGTTG TTGAACCATC TGAAACGGTT GAAAAAGATG GAGATAAGAA ACTCATTATC 1380

TATTAAATGG GGAACTTCAT TTCCAATCAA CGAATTGAAT CTATGGGAGA TGAAGAGAAT 1440

GCTAAGTGGA CTGAACGTGG TGTTCTCATG GATGTCACCA TCAAGAAGAA GGATGGAAAA 1500

ACAACTATCG GAACAGCTAA AGCTCATCCT ACTTGGGTCA ATCGAACACC AAAGGGAACC 1560

TTTTCACCAG AAGGATATCC CTTGTATCAT TACCAAACTT ATATTTTGGA AGATTTTATA 1620

GAGGATGGCA GTCATCGTGA CCAGTTAGAT GAAGCGACTA AGGAACGAAT TGATACAGCC 1680

TATAAAGAAA TGAATGAACA TGTGGGATTG AAGTGGTATT AGCTTGAATC CAGAGGAAAG 1740

TAAATGATGA TTAAGGTAAT TGCGACAGAT ATGGATGGGA CCTTGCTGGA TGCTAGAGGT 1800

CAGCTTGATC TCCCACGATT GGAAAAGATT TTAGATCAGT TGGATCAAAG GGGCATTCGT 1860

TTTGTCATTG CGACGGGCAA TGAAATTCAC CGCATGAGAC AACTACTGAG TCCCTTGGTG 1920

GATCGAGTGG TTCTGGTTGT TGCTAATGGC GCTCGTATTT TTGAAAACAA TGAATTGATT 1980

CAGGCTCAGA CATGGGATGA CGCCATTGTC AACAAGGCTT TGACTCATTT CAAGGGTCGA 2040

GCGTGTCAGG ACCAGTTTGT TGTAACGGGG ATGAAGGGTG ATTTTGTCAA GGAAGGTACG 2100

ATTTTTACAG ATCTTGAAAG TTTTATGACT CCAGAAATGA TTGAAAAATT CTACCAACGG 2160

ATGCAATTTG TGGATGAATT AACATCTGAC CTCTTTGGTG GTGTGCTCAA GATGAGCATG 2220

GTTGTTGGTG AGGAACGTTT GAGTTCGGTT TTGGAAGAAA TCAATGCTCT CTTTGATGGC 2280

CGTGTCCGAG CTGTATCCAG TGGCTATGGT TGCATTGATA TCCTCCAAGC TGGGATTCAT 2340

AAAGCATGGG GCTTGGAGGA ATTACTCAAG CGCTGGGACT TGAAATCCCA AGAAATCATG 2400

GCTTTTGGTG ATAGTGAAAA TGATGTTGAA ATGCTTGAAA TGGCTGGAAT TGCCTATGCG 2460

ATGGAAAATG CTGATGAGAA AGCCAAAGCT GTGGCGACTG CTCTAGCACC AGCCAACAGC 2520

CAAGGAGGAG TTTATCAAGT CTTGGAAAAC TGGTTAGAAA AAGGAGAATG AAGTGGCAGT 2580

ACAGTTATTA GAAAATTGGC TCCTAAAGGA ACAAGAAAAA ATTCAAACTA AGTATCGTCA 2640

CCTAAATCAC ATTTCTGTTG TAGAACCAAA CATTCTTTTT ATTGGGGATT CCATTGTCGA 2700

GTATTATCCT CTACAGGAGC TATTTGGGAC TTCAAAGACG ATTGTCAATC GAGGAATTCG 2760

TGGCTATCAG ACAGGACTGT TACTAGAGAA CCTTGATGCT CATCTATATG GTGGAGCAGT 2820

AGATAAAATT TTTCTTCTGA TTGGGACAAA TGATATCGGA AAGGATGTTC CTGTGAATGA 2880

GGCTCTCAAT AATCTCGAAG CTATCATTCA ATCCGTTGCT CGCGATTATC CATTGACAGA 2940

GATTAAATTG CTTTCCATTT TGCCTGTCAA TGAGAGAGAG GAGTACCAGC AGGCAGTCTA 3000

TATCCGCTCG AATGAAAAAA TTCAGAACTG GAATCAAGCC TATCAAGAGC TTGCATCTGC 3060

CTATATGCAG GTGGAATTTG TGCCAGTATT TGATTGTTTG ACAGACCAAG CAGGCCAACT 3120

CAAAAAAGAA TATACAACTG ATGGACTGCA CCTCAGTATT GCTGGTTATC AGGCTTTGTC 3180

AAAATCCTTG AAAGACTATC TTTACTAAAT AGCTAAATAA TGTTAAATTT GAGCATAATA 3240

TCTTGTAAAA AATTCTAAAA TCCTTTAAAA TAAAAAGTGA CGGAGGAATT TATGAATGTA 3300

AATCAGATTG TACGGATTAT TCCTACTTTA AAAGCTAATA ATAGAAAATT AAATGAAACA 3360

TTTTATATTG AAACCCTTGG AATGAAGGCC TTGTTAGAAG AATCGGCCTT TCTGTCACTA 3420

GGTGACCAAA CGGGTCTTGA AAAGCTGGTT TTAGAAGAAG CTCCCAGTAT GCGTACTCGT 3480

AAGGTAGAGG GAAGAAAAAA ACTAGCTAGA TTGATTGTCA AGGTGGAAAA TCCCTTAGAA 3540

ATTGAAGGAA TCTTATCTAA AACAGATTCG ATTCATCGAT TATATAAAGG TCAAAATGGC 3600

TACGCTTTTG AAATTTTCTC ACCAGAAGAT GATTTGATTT TGATTCATGC GGAAGATGAC 3660

ATAGCAAGTC TAGTAGAAGT AGGAGAAAAG CCTGAATTTC AAACAGATTT GGCATCAATT 3720

TCTTTAAGTA AATTTGAGAT TTCTATGGAA TTACATCTCC CAACTGATAT CGAAAGTTTC 3780

TTGGAATCAT CTGAAATTGG GGCATCCCTT GATTTTATTC CAGCTCAGGG GCAGGATTTG 3840

ACTGTGGACA ATACGGTTAC CTGGGACTTA TCTATGCTCA AGTTCTTGGT CAATGAATTA 3900

GACATAGCAA GTCTTCGCCA GAAGTTTGAG TCTACTGAAT ATTTTATTCC TAAGTCTGAA 3960

AAATTCTTCC TTGGTAAAGA TAGAAATAAT GTTGAATTGT GGTTTGAAGA AGTATGAAGT 4020

GGACCAAGAT TATTAAAAAA ATAGAAGAAC AAATCGAGGC AGGGATTTAT CCCGGAGCCT 4080

CTTTTGCGTA TTTTAAGGAC AATCAATGGA CAGAGTTCTA TTTAGGCCAG AGTGACCCAG 4140

AGCATGGCTT GCAGACTGAG GCAGGACTAG TTTATGACCT AGCTAGTGTC AGCAAGGTTG 4200

TTGGGGTTGG CACAGTTTGT ACCTTCTTGT GGGAAATAGG TCAATTAGAT ATTGATAGAC 4260

TGGTAATAGA TTTTTTACCT GAGAGTGATT ATCCAGACAT CACTATTCGC CAGCTCTTGA 4320

CTCATGCAAC AGACCTTGAT CCTTTTATTC CTAATCGTGA TCTTTTAACA GCCCCTGAAT 4380

TAAAGGAAGC GATGTTTCAT CTCAACAGAC GAAGTCAGCC AGCCTTTCTT TATTCGGATG 4440

TCCATTTTTT GCTGTTGGGC TTTATTTTGG AAAGAATTTT TAATCAAGAT TTGGATGTGA 4500

TTTTAAAGGA TCAAGTCTGG AAACCTTGGG GAATGACGGA AACTAAGTTT GGGCCAGTTG 4560

AGCTTGCTGT TCCAACAGTT AGAGGTGTAG AGGCAGGCAT AGTGCATGAT CCCAAGGCTC 4620

GTCTCCTGGG TAGACATGCT GGGAGTGCTG GTTTATTTTC GACTATAAAG GATTTACAAA 4680

TCTTTTTAGA ACACTATTTA GCAGATGATT TTGCAAGAGA CTTAAATCAA AATTTTTCTC 4740

CTTTGGATGA CAAGGAACGT TCTTTAGCAT GGAATTTGGA AGGAGATTGG CTAGACCATA 4800

CGGGCTATAC AGGTACCTTT ATCATGTGGA ATCGTCAGAA GCAAGAAGCC ACTATTTTCC 4860

TATCGAATCG TACCTATGAA AAGGACGAGA GAGCTCAATG GATATTAGAC CGCAATCAAG 4920

TGATGAACTT GATTCGCAAA GAAGAGTAAG GAGAGACATG TCAAATAGTT TAAAAGGGAC 4980

TTTACTAACA GTTGTGGCTG GTATTGCTTG GGGGTTGTCA GGAACGAGTG GCCAATACCT 5040

AATGGCACAC GGAATTTCGG CTCTGGTCTT GACTAACTTG CGTCTTTTAA TCGCTGGTGG 5100

AATTCTCATG CTCTTGGCTT ATGCTACTGC AAAGGATAAA ATACTGGTCT TTTTAAAGGA 5160

TAGAAAGAGT TTGCTGTCTC TTCTTATTTT TGCTCTGATT GGTCTTTTTC TCAACCAATT 5220

CGCCTATCTG TCTGCTATTC AGGAGACCAA TGCGGGAACA GCGACGGTGC TTCAGTATGT 5280

TTGTCCTGTC GGAATTTTAA TTTATAGCTG TATCAAGGAT AGGGTGGCAC CGACACTGGG 5340

AGAGATAGTT TCCATCATAT TCGCCATCGG AGGAACCTTC CTGATCGCAA CACATGGGCA 5400

GTTGGACCAG TTATCCATGA CACCTGCTGG TCTGTTCTGG GGTCTCTTTT CTGCCTTGAC 5460

TTATGCTCTG TATATCATTT TACCCATAGC CTTGATTAAA AAGTGGGGGA GCAGCTTGGT 5520

CATTGGTGTG GGAATGGTCA TAGCAGGTTT GGTCGCCCTT CCTTTTACAG GGGTTCTACA 5580

GGCCGATATC CCGACTAGTC TTGATTTTCT CCTTGCGTTT GCAGGCATTA TCCTTATCGG 5640

GACTGTCTTT GCCTATACAG CTTTCCTTAA AGGAGCCAGT CTGATAGGAC CGGTCAAGTC 5700

AAGCTTGTTG GCTTCAATTG AGCCAATATC GGCGATTTTC TTTGCCTTCT TAATAATGAA 5760

TGAACAATTT TATCCCATTG ATTTTCTTGG TATGGCAATG ATATTGTTTG CTGTAACTTT 5820

GATTTCTTTG AAAGATTTAT TCTTAGAAAA ATAAAAAAGA CTCTTTGTCC GTGACAGAGA 5880

GTTTTTGCGT GGTAATCTAA TTATTTTCAA GATAAAATTC AAAGCGTTCG CCTACATATT 5940

GACTTTTTAC GTATTCAAAA GCAGTACCAT CTTCTAGGTA GGAAACCTGG GTCAATCCAA 6000

GAATAGCATG TCCTTTTTCA ACTTCCAAAT AGTGGGCAAT CTTTTCTTTA GCAAGGCGAG 6060

CATAGATGGT CTGTTGAGAT TTGCCGATAC GATAGCCATG TTTTTGCAAG GTTTGGAAGA 6120

AATGACTGGT GATTTCTTCT TTTTTAAAGT CCTTAATGAA TTTTTCAGGA ATAGAAGCAA 6180

CTTCATAAAC TAGGGGAACT TGGTCGGCAT AGCGGACCCG CTCCATTCGG ATAATATTGT 6240

CCGTTGGAAA AATTCCTAGC TTGGCAACTT CTTGCTCATT GGGAATGGTT TTTTTGTAGG 6300

AAATGAGCTG GCTAGAGGGA ACTTTACCTT GGGATTTGAC AATTTCAGTA AAACTGGTTG 6360

TCCCTCGCAT CTTTTCTTGT ACTCGAGTAC TGGAAACAAA GGTGCCGCTT CCTACACGGC 6420

GCTCTAAGAC GCCTTCTTCG ACTAATAGAG ATACGGCTTG GCGGAGGGTC ATGCGACTGA 6480

CCGCAAACTG CTCAGCTAAA TCTCTTTCAC TGGGAAGCCT CTCACCAATA GCCCAACGGT 6540

ACTCGTCAAT ATCCTTTTTT ATCTGATCAT GGATTTTTAT ATAAGCAGGT AGCATATTTT 6600

TCACTTCATT TCTATCTTTT CTCTATTGTA CCCCAATAAA CTAGAAAAAG TCAAACTTCG 6660

CCTTGTTTAG TTGGTAATTC GCCCTTATTT GTGATAGAAT ATTGAGAAAA GATATTTCTT 6720

TTGAGAAAGG AAAAAGATGA GCAACATTTC AACTGATTTG CAAGATGTAG AAAAAATCAT 6780

CGTATTGGAC TATGGTAGCC AGTACAACCA GCTGATTTCA CGCCGTATCC GTGAGATTGG 6840

TGTTTTTTCA GAACTAAAAA GCCATAAAAT TTCAGCTGCT GAAGTTCGTG AAGTCAATCC 6900

TGTAGGAATT ATTCTATCAG GTGGTCCAAA TTCTGTATAT GAAGATGGTT CATTTGATAT 6960

TGACCCAGAA ATCTTCGAAC TCGGAATTCC AATTTTGGGA ATCTGTTATG GTATGCAGTT 7020

ATTGACCCAT AAACTTGGAG GAAAAGTTGT TCCTGCAGGT GATGCTGGAA ATCGTGAATA 7080

CGGTCAATCA ACCCTAACTC ACACACCATC AGCGCTTTTT GAATCAACAC CTGATGAACA 7140

GACTGTTTTG ATGAGCCATG GTGATGCGGT TACTGAGATT CCTGCTGACT TTGTTCGTAC 7200

AGGTACATCA GCTGACTGCC CATACGCAGC CATCGAAAAC CCAGATAAAC ACATTTACGG 7260

TATCCAATTC CACCCAGAAG TTCGTCATTC TGTATACGGA AATGATATCC TTCGTAACTT 7320

TGCCCTTAAC ATTTGTAAGG CTAAAGGTGA CTGGTCAATG GATAATTTCA TTGACATGCA 7380

GATCAAAAAA ATTCGTGAAA CCGTCGGTGA TAAACGTGTC CTTCTTGGTC TATCAGGTGG 7440

TGTTGACTCA TCTGTCGTTG GGGTTCTTCT CCAAAAAGCG ATTGGCGATC AATTGATCTG 7500

TATCTTCGTA GACCACGGTC TTCTTCGTAA AGGCGAAGCT GATCAAGTTA TGGACATGCT 7560

CGGTGGTAAG TTTGGTTTGA ATATCGTCAA AGCAGACGCT GCTAAACGTT TCCTTGACAA 7620

ACTTGCTGGC GTTTCTGACC CTGAACAAAA ACGTAAAATC ATCGGTAACG AGTTTGTCTA 7680

TGTATTCGAT GACGAAGCAA GCAAGCTCAA AGATGTGAAA TTCCTTGCTC AAGGTACTTT 7740

ATATACAGAT GTTATCGAGT CTGGTACGGA TACAGCTCAA ACTATCAAGT CACACCACAA 7800

CGTGGtGGTC TTCCAGAAGA TATGCAGTTT GAATTGATTG AACCACTCAA TACTCTTTAC 7860

AAGGATGAAG TTCGTGCTCT TGGTACAGAG CTTGGTATGC CAGACCATAT CGTATGGCGC 7920

CAACCATTCC CAGGACCAGG ACTTGCTATC CGTGTCATGG GTGAAATCAC TGAAGAGAAA 7980

CTTGAAACCG TTCGTGAATC AGACGCTATT CTTCGTGAAG AAATCGCTAA AGCTGGACTT 8040

GACCGCGATA TTTGGCAATA CTTCACTGTT AACACAGGCG TTCGTTCAGT CGGTGTTATG 8100

GGTGACGGTC GTACGTATGA CTACACGATT GCAATCCGTG CTATCACTTC TATCGATGGT 8160

ATGACTGCTG ATTTTGCCAA AATTCCATGG GAAGTACTTC AAAAAATCTC AGTACGTATC 8220

GTAAATGAAG TGGATCATGT TAACCGTATC GTCTACGATA TTACAAGTAA ACCACCTGCA 8280

ACAGTTGAGT GGGAATAATC GCAAAAAAAT TAAAAGCTTT GTAAAATCAA CGGTTACAGA 8340

GGATTAAAAA CTGTAACTGG GATTAAAACG GGAACATTTG CTAAAAAGAA TAAATTGAAT 8400

AATAGTTCCA AGTGGTTTAC ATTTGGACAA AAAATTAGAC CGTAGTTTTC AAGCTGCGGT 8460

CTTTTGATAT ATATAATGAG AATTAATGGC TCTTTGTCAA CTGTAGTGGG TTGAAGTCAG 8520

CTAAGCTCGA GAAAGGACAA ATTTTGTCCT TTCTTTTTTG ATATTCAGAG CGATAAAAAT 8580

CCGTTTTTTG AAGTTTTCAA AGTTCCGAAA ACCAAAGGCA TTGCGCTTGA TAAGTTTGAT 8640

GAGATTATTG GTCGCTTCCA ATTTGGCGTT AGAATAGTGT AGTTGAAGGG CGTTGACGAT 8700

TTTCTCTTTG TCCTTTAGAA AGGTTTTAAA GACAGTCTGA AAAAGAGGAT GAACCTGCTT 8760

TAGATTGTCC TCAATGAGTC CGAAAAATTT CTCCGGTTCC TTATTCTGAA AGTGAAACAG 8820

CAAGAGTTGA TAGAGCTGAT AGTGATGTTT CAAGTCTTGT GAATAGCTCA AAAGCTTGTT 8880

TAAAATCTCT TTATTGGTTA AATGCATACG AAAAGTAGGG CGATAAAAAT GTTTATCGCT 8940

GAGTTTACGA CTATCCTGTT GTATGAGCTT CCAGTAGCGC TTGATAGCCT TGTATTCATG 9000

AGACTTTCGA TCCAATTGAT TCATGATTTG AACACGCACA CGACTCGG 9048

(2) INFORMATION FOR SEQ ID NO: 160:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10399 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

GTACCTTTAT TGATGAATGG ACTGTTTAAA TCAGTAGCAC GCCAACCAGA TATGCTTTCT 60

GAGTTTCGTA GTTTGATGTT TTTAGGTGTT GCCTTTATTG AAGGAACTTT CTTTGTAACT 120

CTTGTCTTCT CATTTATTAT CAAATAAATA CATGGAACGA GAAGAAAAGG GAGGATTTTA 180

GATGGAAGAA AGTATTAATC CAATCATCTC TATTGGTCCT GTTATCTTCA ATCTGACTAT 240

GTTAGCCATG ACTTTGTTGA TTGTGGGAGT TATTTTTGTC TTTATTTATT GGGCAAGCCG 300

CAATATGACC TTGAAACCCA AAGGAAAGCA AAATGTACTT GAGTATGTCT ATGACTTTGT 360

TATTGGATTT ACAGAACCTA ACATTGGTTC GCGCTACATG AAAGATTACT CACTCTTTTT 420

CCTTTGTTTA TTCCTTTTCA TGGTGATTGC CAATAACCTT GGCTTAATGA CAAAGCTTCA 480

AACGATCGAT GGGACTAACT GGTGGAGTTC GCCAACCGCT AATTTACAGT ATGACTTAAC 540

CTTATCTTTT CTTGTCATTT TGTTGACACA TATAGAAAGC GTTCGTCGTC GTGGATTTAA 600

AAAAAGTATA AAATCTTTTA TGAGTCCTGT TTTTGTCATA CCGATGAATA TCTTGGAAGA 660

ATTTACAAAC TTCTTATCTT TGGCTTTGCG GATTTTTGGG AATATCTTTG CAGGAGAGGT 720

CATGACGAGT TTGTTACTTC TTCTTTCCCA CCAAGCTATT TATTGGTATC CAGTAGCCTT 780

TGGAGCTAAT TTGGCTTGGA CTGCATTTTC TGTCTTTATT TCCTGCATCC AAGCTTATGT 840

TTTTACTCTT TTGACATCTG TGTATTTAGG GAATAAGATT AATATTGAAG AGGAATAGAA 900

1038

AGGAGTAACT GATGCACGTA ACAGTAGGTG AATTAATTGG TAATTTTATT TTAATCACTG 960

GCTCTTTTAT TCTTTTGCTA GTCTTGATTA AAAAATTTGC ATGGTCTAAT ATTACAGGCA 1020

TTTTCGAAGA AAGAGCTGAA AAAATTGCTT CAGATATTGA CAGAGCTGAA GAAGCCCGTC 1080

AAAAAGCAGA AGTATTGGCT CAAAAACGCG AAGATGAATT GGCTGGTAGC CGTAAAGAAG 1140

CTAAGACAAT CATTGAAAAT GCAAAGGAAA CAGCTGAGCA AAGTAAGGCT AATATCTTAG 1200

CAGATGCTAA ACTAGAAGCA GGACACTTAA AAGAAAAAGC CAATCAAGAA ATTGCTCAAA 1260

ATAAAGTAGA AGCTTTACAG AGTGTTAAGG GTGAGGTCGC AGATTTGACC ATCAGCTTAG 1320

CTGGTAAAAT CATCTCACAA AACCTTGACA GTCATGCCCA TAAAGCACTC ATTGATCAGT 1380

ATATCGATCA GCTAGGAGAA GCTTAATGGA CAAGAAAACA GTAAAGGTAA TTGAAAAATA 1440

CAGCATGCCT TTTGTCCAAT TGGTACTTGA AAAAGGAGAA GAAGACCGTA TCTTTTCAGA 1500

CTTGACTCAA ATCAAGCAAG TTGTTGAAAA AACAGGTCTG CCTTCTTTTT TAAAACAAGT 1560

GGCAGTAGAC GAGTCGGATA AGGAAAAAAC AATTGCTTTT TTCCAAGATT CTGTGTCGCC 1620

TTTATTACAA AACTTTATCC AGGTTCTGGC CTACAATCAC AGAGCAAATC TTTTTTATGA 1680

TGTGCTTGTA GATTGCTTGA ACCGACTTGA AAAAGAAACA AATCGATTTG AAGTGACGAT 1740

TACGTCTGCT CATCCTCTAA CTGATGAACA GAAGACTCGT TTGCTCCCTT TGATTGAGAA 1800

AAAAATGTCT CTGAAAGTAA GGAGTGTAAA AGAACAAATC GATGAAAGTC TCATTGGTGG 1860

TTTTGTCATT TTTGCCAATC ACAAGACAAT TGATGTGAGT ATTAAACAAC AACTTAAAGT 1920

TGTTAAAGAA AATTTGAAAT AGAAAGTGGT GTTCTTTTGG CAATTAACGC ACAAGAAATC 1980

AGCGCTTTAA TTAAGCAACA AATTGAAAAT TTCAAACCCA ATTTTGATGT GACTGAAACA 2040

GGTGTTGTAA CCTATATCGG GGACGGTATC GCGCGTGCTC ACGGCCTTGA AAATGTCATG 2100

AGTGGAGAGT TGTTGAATTT TGAAAACGGC TCTTATGGTA TGGCTCAAAA CTTGGAGTCA 2160

ACAGACGTTG GTATTATCAT CCTAGGTGAC TTTACAGATA TCCGTGAAGG CGATACAATC 2220

CGCCGTACAG GGAAAATCAT GGAAGTCCCT GTAGGTGAAA GTCTGATTGG TCGTGTTGTG 2280

GATCCGCTTG GTCGTCCAGT TGACGGTCTT GGAGAAATCC ACACTGATAA AACTCGTCCA 2340

GTAGAAGCAC CAGCTCCTGG TGTTATGCAA CGTAAGTCTG TTTCAGAACC ATTGCAAACT 2400

GGTTTGAAAG CTATTGACGC CCTTGTACCG ATTGGTCGTG GTCAACGTGA GTTGATTATC 2460

GGTGACCGTC AGACAGGGAA AACAACCATT GCGATTGATA CAATCTTGAA CCAAAAAGAT 2520

CAAGATATGA TCTGTATCTA CGTCGCGATT GGACAAAAAG AATCAACAGT TCGTACGCAA 2580

GTAGAAACAC TTCGTCAGTA CGGTGCCTTG GACTACACAA TCGTTGTGAC AGCCTCTGCT 2640

TCACAACCAT CTCCATTGCT CTTCCTAGCT CCTTATGCTG GGGTTGCTAT GGCGGAAGAA 2700

TTTATGTATC AAGGTAAGCA TGTTTTGATT GTATACGATG ATCTATCAAA ACAAGCGGTA 2760

GCTTATCGTG AACTGTCGCT CTTGCTTCGT CGTCCTCCAG GTCGTGAAGC CTTCCCAGGG 2820

GATGTTTTCT ATCTCCACAG CCGTTTGCTT GAGCGCTCAG CTAAAGTTTC TGATGAACTT 2880

GGTGGTGGAT CAATTACAGC CCTACCATTT ATCGAGACAC AAGCAGGAGA TATCTCAGCC 2940

TATATCGCAA CCAACGTGAT TTCTATCACT GATGGACAAA TCTTCCTTGG CGATGGCCTC 3000

TTCAATGCAG GTATTCGTCC AGCCATCGAT GCGGGTTCAT CTGTATCTCG TGTAGGTGGT 3060

TCTGCACAAA TCAAAGCCAT GAAGAAGGTT GCTGGTACAC TTCGTATCGA CCTTGCTTCA 3120

TACCGTGAGT TGGAAGCCTT TACTAAGTTT GGTTCTGACT TGGACGCAGC AACACAGGCT 3180

AAGTTGAACC GTGGACGTCG TACCGTTGAG GTCTTGAAAC AACCTGTTCA CAAACCATTA 3240

CCTGTTGAGA AACAAGTAAC CATTCTTTAT GCTTTGACAC ATGGTTTCTT GGATACTGTT 3300

CCAGTAGATG ATATTGTTCG TTTCGAGGAA GAGTTCCATG CCTTCTTTGA TGCTCAACAT 3360

CCAGAGATTT TGGAAACCAT TCGTGATACA AAAGACTTGC CAGAAGAAGC AGTCTTGGAT 3420

GCTGCGATTA CAGAGTTTCT CAATCAATCT AGCTTCCAAT AAGAATAGAG GTGTCAGATG 3480

GCAGTATCTC TAAATGATAT TAAAACAAAA ATCGCCTCAA CAAAAAATAC GAGTCAAATC 3540

ACTAATGCCA TGCAAATGGT ATCGGCTGCT AAGCTAGGTC GTTCTGAAGA AGCTGCTCGC 3600

AACTTCCAAG TTTACGCTCA GAAAGTGCGT AAACTTTTGA CAGATATCCT TCATGGTAAT 3660

GGAGCTGGTG CTTCAACTAA TCCGATGTTG ATTAGCCGTT CTGTGAAGAA GACAGGCTAT 3720

ATCGTTATCA CTTCAGACCG CGGTTTGGTT GGAGGTTATA ATTCCTCTAT TTTGAAAGCT 3780

GTTATGGAGT TGAAAGAAGA ATACCACCCA GACGGTAAAG GTTTTGAAAT GATCTGTATC 3840

GGTGGGATGG GAGCTGATTT CTTTAAGGCT CGCGGTATTC AACCACTTTA TGAATTACGT 3900

GGCTTGTCAG ACCAACCTAG CTTTGATCAA GTTCGTAAGA TTATTTCAAA AACTGTTGAA 3960

ATGTACCAAA ATGAACTCTT TGATGAGCTT TATGTTTGCT ACAACCACCA TGTCAATACG 4020

CTAACCAGTC AAATGCGTGT GGAACAAATG CTTCCGATTG TTGACTTGGA TCCAAATGAA 4080

GCGGATGAAG AGTACAGCTT GACTTTTGAA TTGGAAACCA GCCGAGAAGA AATTCTGGAG 4140

CAGTTGTTGC CTCAGTTTGC AGAAAGTATG ATTTACGGTG CCATTATCGA TGCCAAGACA 4200

GCTGAGAATG CTGCGGGCAT GACAGCCATG CAAACAGCGA CAGATAATGC TAAGAAAGTC 4260

ATCAATGATT TGACAATTCA GTATAACCGT GCCAGACAGG CGGCGATTAC ACAAGAAATT 4320

ACAGAAATCG TAGCAGGTGC TAGTGCCTTA GAATAGGCTC TAGTCCAGCT CGTATGAAAA 4380

TGAACTTAGG ACCTAGTTGA GCTAGGAACC GACAGTATCT TATATAGAAT AGGAGAAGGA 4440

1040

GATGAGTTCA GGTAAAATTG CTCAGGTTAT CGGTCCCGTT GTAGACGTTT TGTTTGCAGC 4500

AGGGGAAAAA CTTCCTGAGA TTAACAATGC ACTTGTCGTC TACAAAAATG ACGAAAGAAA 4560

AACAAAAATC GTCCTTGAAG TAGCCTTGGA GTTAGGAGAT GGTATGGTTC GTACTATCGC 4620

CATGGAATCA ACAGATGGGT TGACTCGTGG AATGGAAGTA TTGGACACAG GTCGTCCAAT 4680

CTCTGTACCA GTAGGTAAAG AAACTTTGGG ACGTGTCTTC AACGTTTTGG GAGATACCAT 4740

TGACTTGGAA GCTCCTTTTA CAGAAGACGC AGAGCGTCAG CCAATTCATA AAAAAGCTCC 4800

AACTTTTGAT GAGTTGTCTA CCTCTTCTGA AATCCTTGAA ACAGGGATCA AGGTTATTGA 4860

CCTTCTTGCC CCTTACCTTA AAGGTGGTAA AGTTGGACTT TTCGGTGGTG CCGGAGTTGG 4920

TAAAACTGTC TTAATCCAAG AATTGATTCA CAACATTGCC CAAGAGCACG GTGGTATTTC 4980

AGTATTTGCT GGTGTTGGGG AACGTACTCG TGAGGGGAAT GACCTTTACT GGGAAATGAA 5040

AGAATCAGGC GTTATCGAGA AAACAGCCAT GGTCTTTGGT CAGATGAATG AGCCACCAGG 5100

AGCACGTATG CGTGTTGCCC TTACTGGTTT GACAATCGCT GAATACTTCC GTGATGTGGA 5160

AGGCCAAGAC GTGCTTCTCT TTATCGATAA TATCTTCCGT TTCACTCAGG CTGGTTCAGA 5220

AGTATCTGCC CTTTTGGGTC GTATGCCATC AGCCGTTGGT TACCAACCAA CACTTGCTAC 5280

GGAAATGGGT CAATTGCAAG AACGTATCAC ATCAACCAAG AAGGGTTCTG TAACCTCTAT 5340

CCAGGCTATC TATGTGCCAG CGGATGACTA TACTGACCCA GCGCCAGCAA CAGCCTTCGC 5400

TCACTTGGAT TCAACAACAA ACTTGGAACG TAAGTTGGTA CAATTGGGTA TCTACCCAGC 5460

CGTTGACCCA CTTGCTTCAA GCTCACGTGC CTTGGCACCT GAAATCGTTG GAGAAGAGCA 5520

CTATGCAGTT GCTGCTGAAG TAAAACGTGT CCTTCAACGT TACCATGAAT TGCAAGATAT 5580

CATTGCTATC CTTGGTATGG ATGAGCTTTC TGATGAAGAA AAGACCTTGG TTGCTCGCGC 5640

CCGTCGTATC CAGTTCTTCT TGTCACAAAA CTTCAACGTT GCGGAACAAT TTACTGGTCA 5700

GCCAGGTTCT TATGTTCCAG TTGCTGAAAC TGTACGTGGC TTTAAGGAAA TCCTTGATGG 5760

TAAATACGAC CACTTGCCAG AAGATGCCTT CCGTGGTGTA GGTTCTATCG AAGATGTGAT 5820

TGCAAAAGCT GAAAAAATGG GATTTTAAGA GGTGATCTAT GGCTCAGTTA ACTGTCCAGA 5880

TCGTGACACC AGATGGTCTC GTCTATGATC ACCATGCCAG CTATGTATCG GTTCGAACTC 5940

TGGATGGTGA GATGGGGATC TTGCCACGAC ATGAAAATAT GATTGCGGTT TTAGCAGTTG 6000

ATGAAGTAAA GGTAAAACGT ATCGATGATA AAGATCACGT GAACTGGATT GCAGTAAACG 6060

GTGGCGTTAT TGAAATTGCC AATGATATGA TCACAATCGT CGCTGACTCT GCAGAACGTG 6120

CTCGTGATAT CGATATCAGT CGTGCAGAAC GTGCCAAACT TCGTGCAGAA CGTGCAATTG 6180

AAGAAGCACA AGACAAACAT TTGATTGACC AAGAACGTCG TGCTAAGATT GCTTTGCAAC 6240

GTGCTATTAA CCGTATTAAT GTCGGAAATA GACTATAAGA AAAAATGAAC TTGAAAATAC 6300

CAAGTTCATT TTTTATGGTG TTTTAAGGAG CAAAACGGAT GCAGACTGCT TCGGGAACAT 6360

GGAAGTCGTT GGAGAGTTCT GCTAGACGAC CATTGTCACA ATTACGTTTA AAGACAGTTG 6420

CATTGTCAGA GTCTTGATGG ACAACAATGA GAAATTTTTG GTCGGGTGTC AAATCAAAAT 6480

CACGTGGAGT CTGACCATGC GTTGGAACGA TTTCTAATAA CTCTAAGCTA CCGTCCGCAA 6540

GGATGGTATA TACTGCGATA GAATCATGGC CACGGTTAGA AGCGTAGAGG TATTTACCGT 6600

CTTTAGAGAG ATGAATAGCA GCGGTTCCAT TAAAGCCTTC GTAAGCTTCC GGTAAAGTTG 6660

AAATGACCTG CATACGTTCA AATTCGCCAA CGCCATCGTA GATTAAAACT TCGATAGTAC 6720

TATTGAGTTC ACAAATGAGA TAAGCGATTT TATAGTGGTT ATGGAAAATG ATATGGCGTG 6780

AGCCTGCTCC TGGCTTGCTG TGATAGGTAT AGAGCTTAGA TAATTTTCCT TCTTGATCGA 6840

GGTCATAGGT GATGACTTGG TCAGTTCCCA AGTCGCAGGT CACTAGATAG TGGTCAGGTG 6900

TTAAATCTGT ATAGTGAACA TGGGGGGAAG CTTGATTTTC ATGTGGACCT TGGCCACTGT 6960

GTTGATCCAT ATCACTAAGT AGAAGACTAC CATCTTCCTG GCGTTTATAA ACAAGGACTT 7020

GTCCCTTGTG ATAGTTAGCT GCGTAAACCA AATCACGCTT TTCATCGACA GCAACATAAC 7080

AGTGGGGAGC TCCTTCTTCA ACAACATGAT TTAACACAGT CCCGTCAGTT TGATAGGCTG 7140

CAATTCCCCC CTTATCGTCT TGGCTACCAA CAGTGTATAA ATGTTGGTGC TGGTCAAAGG 7200

CAAGGTAGGT TGGACTTGGC TCAGCTGCAA AAAGTTCTAG ATTTGAAAGC TGACCAGTTT 7260

CTGTATCAAA GTCTGCCTTG TAAATCCCTT GAGAAGTACG ACGTGTATAA GTTCCAAAAT 7320

AAACAGTTTC TTTCATTACT ATACCTCTGT GTAAAGATAA GACTATTATA TCACAAAAAC 7380

AAGTAAATTA AAGATATCCA ATTAGATGTA AGCACTTTAA AAAAGAGTTA TTTTGTTTCA 7440

AAAATGGTAT AATGAGAGAA CAATAGAAAG GAAGTATTTA TGGAGCAAAA AGAGAAACAT 7500

TTTAGCCTAT CTTGGTTTTT CAAGTGGTTT TTAGATAACA AGGCAATTAC GGTATTTTTA 7560

GTAACCTTAT TATTGGGACT GAATCTTTTT ATTTTAAGTA AGATTAGTTT TCTATTTTCA 7620

CCTGTTTTAG ACTTTTTAGC AGTTGTGATG TTGCCAGTCA TTTTGTCTGG TTTGTTATAT 7680

TATTTGTTGA ATCCTATTGT TGATTGGATG GAGAAGCATA AGGTTAATCG TGTTATAGCT 7740

ATCACTATTG TCTTTGTTAT CATCGCTCTC TTTATCATTT GGGGCTTGGC AGTCGCCATT 7800

CCAAATCTGC AACGTCAGGT TTTGACCTTT GCAAGAAACG TTCCTGTTTA CTTAGAAGAT 7860

ATAGATAGGA TTGTTAATGG ATTGGTAGCC CAGCACCTGC CAGATGATTT CAGACCTCAA 7920

TTAGAGCAAG TTTTGACCAA TTTTTCTAGC CAGGCTACAG TTTTGGCAAG TAAGGTTTCA 7980

TCTCAGGCAG TCAACTGGGT GAGTGCCTTT ATTAGCGGGG CTTCTCAAGT GATTGTTGCC 8040

TTGATTATCG TTCCTTTCAT GCTCTTTTAT CTCTTGCGTG ATGGGAAAGG CTTGCGTAAC 8100

TATTTGACCC AATTCATTCC AAGAAAATTG AAGGAACCTG TTGGACAAGT TTTATCAGAT 8160

GTGAATCAAC AGTTGTCCAA CTATGTTCGA GGGCAAGTGA CAGTGGCTAT TATTGTAGCA 8220

GTAATGTTTA TCATCTTCTT CAAGATTATT GGTCTACGCT ATGCGGTTAC GCTGGGGGTT 8280

ACTGCTGGTA TTTTAAATCT GGTCCCTTAT CTTGGTAGCT TTCTAGCCAT GCTTCCTGCT 8340

CTAGTATTGG GTTTGATTGC TGGTCCAGTC ATGCTTTTGA AAGTAGTGAT TGTCTTTATC 8400

GTAGAACAAA CTATTGAAGG CCGTTTTGTC TCTCCATTGA TTTTGGGAAG TCAATTAAAC 8460

ATCCACCCTA TTAATGTTCT CTTTGTTTTG TTAACTTCAG GATCTATGTT TGGTATCTGG 8520

GGAGTTTTAC TTGGTATTCC GGTTTATGCC TCTGCTAAGG TTGTCATTTC AGCCATTTTC 8580

GAATGGTATA AGGTAGTCAG TGGTCTATAT GAATTAGAGG GTGAGGAAGT CAAGAGTGAA 8640

CAATAGTCAA CAGATGTTAC AGGCTTTGGA GGAGCAAGAT TTAACTAAGG CTGAGCATTA 8700

TTTCGCCAAA GCTTTAGAAA ATGATTCAAG TGATCTTCTG TATGAATTGG CAACTTATCT 8760

TGAAGGGATT GGTTTCTATC CTCAGGCCAA GGAAATTTAC CTGAAAATTG TAGAGGATTT 8820

TCCAGAGGTT CATCTTAATC TAGCTGCAAT TGCTAGCGAG GATGGTCAAA TAGAAGAAGC 8880

CTTTACCTAT CTTGAGGAAA TCCAAGCTGA CAGTGACTGG TATGTCTCGT CTTTGGCTCT 8940

GAAGGCAGAC CTTTACCAGC TGGAAGGTTT GACAGATGTG GCACGTGAGA AATTATTGGA 9000

GGCCTTGACC TACTCAGAGG ATTCTCTCTT GATATTGGGT TTGGCAGAGT TGGATAGTGA 9060

GTTGGAAAAT TACCAAGCGG CTATTCAAGC CTATGCCCAG TTAGATAATC GCTCGATTTA 9120

TGAGCAAACG GGCATTTCCA CCTATCAACG AATTGGCTTT GCCTATGCTC AGTTAGGGAA 9180

ATTTGAAACG GCTACTGAGT TTTTAGAAAA AGCCCTGGAG TTAGAATACG ATGACTTAAC 9240

AGCTTTTGAG TTGGCCAGTC TTTATTTTGA TCAAGAAGAA TATCAAAAAG CCACCCTCTA 9300

CTTTAAGCAG CTTGATACCA TTTCTCCTGA CTTTGAAGGC TATGAGTATG GGTACAGTCA 9360

GGCTTTACAT AAGGAACATC AAGTTCAAGA AGCCCTGCGT ATCGCTAAGC AAGGATTAGA 9420

GAAAAATCCC TTTGAAACTC GCCTCTTGCT AGCTGCTTCA CAATTTTCTT ATGAATTGCA 9480

TGATGCTAGT GGTGCAGAAA ATTATCTCCT TACTGCAAAA GAAGACGCTG AGGATACAGA 9540

AGAAATCTTG CTTCGTTTAG CCACTATTTA TCTGGAGCAG GAGCGTTATG AGGATATTCT 9600

AGAATTGCAG AGTGAGGAGC CAGAAAATCT TTTGACCAAG TGGATGATTG CTCGTTCTTA 9660

TCAAGAAATG GACGATTTGG ATACTGCTTA TGAGTATTAT CAAGAGTTGA CAGGAGATTT 9720

GAAGGACAAT CCAGAATTTC TGGAACACTA TATCTATCTC TTGCGTGAAT TGGGACATTT 9780

TGAAGAAGCA AAAGTCCATG CTCACACTTA CTTAAAACTG GTTCCAGATG ATGTGCAAAT 9840

GCAAGAACTG TTTGAGAGAT TGTAAGAATG TTTAACCCAA ATCATTCATA CCTCTCTCAA 9900

CTAGATGTAA CTTACAAAAC CCCTGACCTC ATGAGCCACT TTCTTCCTCC TCATGAGGTC 9960

AGTTTTACTT TCTGCTGTTC CAGTATCGTT TTTCCTCGCT AGATTTCCTC AAAAGGGCAG 10020

ACTCCTCCCT TGGTGCGTCA CACGATTTTT TCATCTCGAC TGTTCTTTAA TGCATCATTA 10080

ACGACGCTTT TCTTCTAGGT GGTTCATAAG GAACAGGAAG ATTCAGGTTG ACTTTTCTAA 10140

TCCTAGAATA AAGTGCTGAA AACAATTCGG AATAGGCATA GAGACTAGAC AATTTGAGGA 10200

GCTGCTTGCG TCCTGTTCGA ACACATTTTC CCACCACGTG AAGAAAAAGA TGGCGGAAGC 10260

GTTTGATTGT TAAAGTTTGG AAGTCACCTC CAGCTAGATG TTTGAGAAAA AGATAGAGAT 10320

TGTAGGCGAT ACAGCTCATC ATCATACGAA TTCGTTTTTG ATTAAGGTTG AACTATCCGT 10380

TTTATCGCCA AAAAATCGG 10399

(2) INFORMATION FOR SEQ ID NO: 161:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9409 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

GATAAGATTA AGTTAGAAAA GAAAGAACTA GGACATATCT ACCAGATTCA GGTTTTTAAT 60

AGCTATGGGC AGGAAGAAAT CTATCGTGTG ATTTTGATGG AGACCAATAT TAGTTCGGTT 120

TCAACCAATA TCAAGTATGC TGCTGTCTTG ATTAATACCA GTCAGTTGGA ACAGGCTAGT 180

CAAAAGCATG AGCAATTGAT TGTGGTCGTG ATGGCTAGTT TCTGGATTTT GTCTTTACTT 240

GCCAGTCTCT ATCTAGCTAG GGTCAGTGTT AGGCCCCTGC TTGAGAGTAT GCAGAAGCAA 300

CAGTCTTTTG TGGAAAATGC CAGTCATGAG TTACGAACTC CACTCGCAGT TTTGCAAAAT 360

CGCTTAGAGA CCCTTTTTCG TAAGCCAGAA GCTACCATTA TGGATGTGAG CGAAAGCATT 420

GCATCGAGTT TGGAAGAAGT CCGAAATATG CGTTTTTTAA CGACAAGCTT GCTGAACTTA 480

GCTCGGAGAG ATGATGGGAT TAAGCCGGAG CTTGCAGAAG TTCCAACTAG CTTTTTTAAT 540

ACAACTTTCA CAAACTACGA GATGATTGCT TCGGAAAATA ATCGTGTCTT CCGTTTTGAA 600

AATCGTATCC ATCGAACAAT TGTCACAGAT CAGCTTCTTC TGAAACAACT GATGACCATT 660

CTTTTCGATA ATGCCGTCAA GTATACTGAG GAGGATGGTG AAATTGATTT TCTTATCTCG 720

GCGACCGATC GCAATCTTTA TTTACTTGTT TCTGATAATG GAATCGGTAT TTCGACAGAA 780

GATAAAAAGA AAATTTTTGA CCGTTTTTAT CGAGTAGACA AGGCTAGAAC CCGGCAAAAA 840

GGTGGTTTTG GTTTAGGATT ATCCCTAGCC AAGCAAATTG TAGATGCTCT AAAAGGAACT 900

GTTACTGTCA AAGATAATAA ACCCAAGGGA ACAATCTTTG AAGTGAAGAT TGCCATTCAG 960

ACACCATCTA AAAAGAAAAA ATAAAAATAT CGCTCCAATT GGGGCGATAT TTTGGATTTA 1020

TCTTCTACGT TTTCGTTTGA TAATAGACCG TTGAACTTTT AAAACAAGTA AGCTGAATCC 1080

GATTGCTGCG GCAAAGGCAA GAGCAGTTGA TAATTTTAAT GCTAAAAAGA TAAAACTAAA 1140

GATAGCAATA CAGATACAAA AAACAGCGAT ATTAATAAAA AATAGGATTT CCTTGAGATT 1200

GGCATCAGAT TGCGCTTCAG GTGTATAAGC TTGGTAATGA GGAAGCTGCT GGTTTAATTC 1260

TTCTTGATAG TCTACCTCAT AGGATTGTAA TTTTCTTACG GGCATGATTC TCTCCTTAAC 1320

AGTACATACC TATTTTATCA TTTTTTCGGC AGAGAATTAT TACAGAAAGG TTACAAAAAG 1380

AATAAAGTCC CTTTTCATTT TCAAAGCATG GCTGATTTTG GAGAAATGTG GTATAATTTT 1440

TCTTATGGAA AAGATTGTCA TTACAGCAAC TGCTGAAAGT ATTGAACAAG TTGAACAACT 1500

ACTCGAAGCT GGCGTAGACC GTATCTATGT CGGTGAGAAA GATTTTGGTC TTCGTCTGCC 1560

AACGACCTTT AGTTATGACC AATTACGTGA AATCGCTAAG TTGGTTCATG ATGCTGGTAA 1620

GGAATTGATC GTTGCGGTCA ATGCTCTCAT GCACCAAGAT ATGATGGACC GTATCAAGCC 1680

TTTCTTAAAC TTCTTGGAAG AAATCAAGAC AGACTATATT ACGATTGGGG ATGCAGGCGT 1740

CTTTTACGTA GTTAACCGCG ATGGTTATTC ATTTAAGACC ATCTACGATG CTTCAACCAT 1800

GGTAACTAGC AGTCGTCAGA TTAACTTCTG GGGACAAAAG GCTGGCGCAT CTGAGGCTGT 1860

TTTGGCGCGT GAAATTCCAT CAGCTGAACT TTTCAAAATG CCAGAGATTT TGGAAATTCC 1920

TGCTGAAGTT TTGGTTTACG GTGCTAGCGT CATCCATCAT TCTAAACGTC CACTCTTGCA 1980

AAACTACTAT AACTTTACAC ATATCGATGA TGAAAAGACG CATAAACGTG ACCTCTTCTT 2040

GGCTGAGCCA AGTGATCCAG AGAGCCACTA TTCCATTTTT GAAGATAATC ATGGGACCCA 2100

TATCTTTGCC AACAATGACC TTGATTTGAT GATCAAATTA ACAGAATTGG TGGAGCATGG 2160

CTTTACTCGC TGGAAACTAG AAGGGCTCTA CACTCCTGGT CAGAACTTTG TTGAGATTGC 2220

AAAACTCTTT ATCCAAGCGC GTAGCTTGAT TCAAGAGGGC AACTTTAGTC ATGCTCAAGC 2280

CTTCTTGCTG GATGAAGAAG TTCGTAAACT TCACCCTAAA AACCGTTTCC TTGATACAGG 2340

ATTTTATGAC TACGATCCTG ACATGGTTAG ATAAAATACA TGATTCGTTG AGAGAAGGAA 2400

GATGCAAACA TTTCTTCTCT CAATTTTTCG TATTTCTTCA CTATTTTACA AAAATCAGCA 2460

GGCTAGAATG CTCTATTCGA TGGGATTTTT AAGAAAAGTA GTGTTCTTGA GTTTGAAAAT 2520

TATCCTATGT TTGCAGGTGC CAAATGGCCC TTTTTTTGGT ATAATTTTTT ATAATGAAAA 2580

CGATTGGTAA TCGCTATGTT GTGGTGGATT TAGAGGCAAC TAGCACAGGT AGTAAGGCTA 2640

AAATTATCCA AGTGGGAATT GTCGTGATTG AGGACGGAGA AATCGTCGAT CACTATACGA 2700

CGGATGTCAA TCCACATGAA CCCTTGGATG CTCATATCAA AGAACTGACA GGATTGACAG 2760

ACCAACGTCT GGCGCAAGCA CCTGATTTTT CGCAAGTTGC CAGAAAAATA TTTGACTTGG 2820

TGGAGGATGG GATTTTTGTA GCCCATAATG TTCAGTTTGA TGCTAATCTC TTGGCGGAAA 2880

ATTTATTTTT TGAAGGCTAT GAGCTAAGAA ACCCTCGTGT TGATACGGTC GAATTGGCCC 2940

AGGTCTTTTT CCCTGAACTG GAAAAATATA GCTTGCCGAT TTTGTGTCGA GAATTAGGAA 3000

TTCCTCTTAA ACACGCACAC ACAGCCCTTT CAGATGCCCA AGCTACAGCA GAATTACTTC 3060

TTTTTTTACG GAAAAAGATG ACCCAGCTTC CTAAAGGTCT CTTGGAACGC TTGCTGGAAA 3120

TGGCTGACGC TCTCCTATAT GAGTCCTACC TGGTTATTGA GGAAACTTAT CGCAACCAAT 3180

CTATCCTGAG TTCTCCAGAC TTGGTCCAAG TTCAAGGTCT ATATTTTAAG AAAACGGAAG 3240

CTTCTCTGGA GCCACGAAAA CTATCTCAAG ACTTTTCTAA AAATATTTCT CTGTTGAACC 3300

TTGAAGTGAG GGAGGAACAA GAAAGTTTTG CTAAAGAGGT TGGCTTGCTA TTGAAAGATG 3360

AACCTGTCTC TCTGATTCAA GCGCCGACAG GGATTGGGAA AACCTATGGC TATCTCTTAC 3420

CCGCTTTATC TCAATCCAAA GAGCGACAAA TTGTTCTTAG TGTTCCGACA AAGATTCTTC 3480

AAAATCAAAT CATGGAAGAA GAAGGTAAAC GCCTCAAGGA AGTGTTCCAT ACAGATATTC 3540

ATAGCTTAAA GGGACCACAA AATTATCTGA AGTTGGATGC CTTTTATCAT TCCTTGCAGG 3600

AAAATGATGA AAATCGCTTA TTTAGACGCT TTAAAATGCA AGTCTTGGTC TGGCTTACTG 3660

AGACAGAGAC AGGAGATTTG GATGAAATCG GGCAACTCTA CCGTTACCAA CATTTTCTAG 3720

CAGACCTTCG TCATGATGGG AATTTATCAT CCCAGAGCTT ATTTGTGACG GAAGATTTTT 3780

GGAAACGTAG TCAAGAAAGG GCAGAGACTT GCAAGCTTTT AGTGACTAAT CATGCCTATC 3840

TCGTAACCAG ACTTGAAGAT AATCCTGAAT TTGTCAGTGA CCGTTTACTG ATTATTGATG 3900

AAGTCCAAAA GATTTTGTTA GCTCTAGAAA ATCTGCTTCA AGAGACCTAC GATATACAAT 3960

CTATTATCGA TTTAATTGAT AAGGCTTTAG TAGGAGAAGA AAACAGGGTT CAACAACGGA 4020

TACTAGAAAG TATTCGCTTT GAGTGTCTCT ACTTGATAGA ACAATTTCAG TCTGGCAAAT 4080

CTAGGAAAAA TATCTTAGAT TCTCTGGACA ATCTCCATCA GTATTTTTCA GAATTGGAAG 4140

TAGAAGACTT TGATGAGCTG GTTCGCTATT TTACAGCTGA AGGTGATTAC TGGCTTGAAG 4200

TAACTGAAAC GAGTCAAAAG AAAATTCAGA TTTCTTCTAC AAAATCAGGC CGTACTCTTC 4260

TGTCCTCTTT ACTTCCTGAG AGTTGCCAAG TCTTGGGAGT ATCGGCTACT CTTGAGATTA 4320

GTCAGAGGGT TTCTTTGGCA GACCTTTTAG GCTATCCTGA AGCTAAATTT GTCAAGATTG 4380

AATCTCGGGG AAAACAGGAA CAAGAAGTGG TCATGGTCAA AGATTTCCCT CTGGTAACAG 4440

AAACCTCCTT AGAAGTCTAT GCCAGAGAGG TAGCTGCTTT ACTAGTGGAA ATTCAAGCTT 4500

TCCAGCAACC GATTTTGGTT CTCTTTACCG CTAAAGACAT GCTTCTAGCA GTATCGGATT 4560

TACTTACAGT TAGCCACTTG GCCCAGTATA AAAATGGGGA TGTTCATCAG CTAAAGAAAC 4620

GCTTTGAAAA AGGTGAACAA CAAATCTTGC TTGGTGCAGC AAGTTTCTGG GAGGGAGTTG 4680

ATTTTTCAAG CCATCCTTCT GTGATTCAAG TTGTACCGAG GCTTCCTTTC CAAAATCCTC 4740

AAGAACCCTT GACGAAAAAG ATTAATCAAG AACTGAATCA AGAAGGGAAA AATGCCTTTT 4800

ATGATTATCA ATTGCCAATG GCCATTATTC GTTTAAAACA GGCTTTGGGA AGAAGTATGA 4860

GACGTGAATA CCAACGTTCC TTAACTCTTA TTTTGGATAG GAGAATCGTC GGAAAACGAT 4920

ACGGCAAACA AATAGTAGCA TCTCTAGCAG AAGAAGCGAC TGTTAAAACC ATCTCTCGAT 4980

CCGAAGTTGA CGAGGCTATT GATAGATTTT TTAATGAGCT TTGATAAATA GTATTGTATG 5040

AAAGTATAAG GTTAGTATAT ATGAAACGTT CTCTCGACTC AAGAGTCGAT TACAGTTTGC 5100

TCTTGCCAGT ATTTTTTCTA CTGGTCATCG GTGTGGTGGC TATCTATATA GCCGTTAGTC 5160

ATGATTATCC CAATAATATT CTGCCCATTT TAGGGCAGCA GGTCGCCTGG ATTGCCTTGG 5220

GGCTTGTGAT TGGTTTTGTG GTCATGCTCT TTAATACAGA ATTTCTTTGG AAGGTGACCC 5280

CCTTTCTATA TATTTTAGGC TTGGGACTTA TGATCTTGCC GATTGTATTT TATAATCCAA 5340

GCTTAGTTGC ATCAACGGGT GCCAAAAACT GGGTATCAAT AAATGGAATT ACCCTATTCC 5400

AACCGTCAGA ATTTATGAAG ATATCCTATA TCCTCATGTT GGCTCGTGTC ATTGTCCAAT 5460

TTACAAAGAA ACATAAGGAA TGGAGACGCA CGGTTCCGCT GGACTTTTTG TTAATTTTCT 5520

GGATGATTCT CTTTACCATT CCAGTCCTAG TTCTTTTAGC ACTTCAAAGT GACTTGGGGA 5580

CGGCTTTGGT TTTTGTAGCC ATTTTCTCAG GAATCGTTTT ATTATCAGGG GTTTCTTGGA 5640

AAATTATTAT CCCAGTATTT GTGACTGCTG TAACAGGAGT TGCTGGTTTC TTAGCTATCT 5700

TTATTAGCAA GGACGGACGA GCTTTTCTTC ACCAGATTGG AATGCCGACC TACCAAATTA 5760

ATCGGATTTT GGCTTGGCTC AATCCCTTTG AGTTTGCCCA AACAACGACT TACCAGCAGG 5820

CTCAAGGGCA GATTGCCATT GGGAGTGGTG GCTTATTTGG TCAGGGATTT AATGCTTCGA 5880

ATCTGCTTAT CCCAGTTCGA GAGTCAGATA TGATTTTTAC GGTTATTGCA GAAGATTTTG 5940

GCTTTATTGG CTCTGTCCTG GTTATTGCCC TCTATCTCAT GTTGATTTAC CGTATGTTGA 6000

AGATTACTCT TAAATCAAAT AACCAGTTCT ACACTTATAT TTCCACAGGT TTGATTATGA 6060

TGTTGCTCTT CCACATCTTT GAGAATATCG GTGCTGTGAC TGGACTACTT CCTTTGACGG 6120

GGATTCCCTT GCCTTTCATT TCGCAAGGGG GATCAGCTAT TATCAGTAAT CTGATTGGTG 6180

TTGGTTTGCT TTTATCGATG AGTTACCAGA CTAATCTAGC TGAAGAAAAG AGCGGAAAAG 6240

TCCCATTCAA ACGGAAAAAG GTTGTATTAA AACAAATTAA ATAAGGAGAA AATCATGGTA 6300

AAAGTAGCAG TTATATTAGC TCAGGGCTTT GAAGAAATTG AAGCCTTGAC AGTTGTAGAT 6360

GTCTTGCGTC GAGCCAATAT CACATGTGAT ATGGTTGGTT TTGAAGAGCA AGTAACGGGT 6420

TCGCATGCAA TCCAAGTAAG AGCAGATCAT GTCTTTGATG GAGATTTATC AGACTATGAT 6480

ATGATTGTTC TTCCTGGAGG TATGCCTGGT TCTGCACATT TACGTGATAA TCAGACCTTG 6540

ATTCAAGAAT TGCAAAGCTT CGAGCAAGAA GGGAAGAAAC TAGCAGCCAT TTGTGCGGCA 6600

CCAATTGCCC TCAATCAAGC AGAGATATTG AAAAATAAGC GATACACTTG TTATGACGGC 6660

GTTCAAGAGC AAATCCTTGA TGGTCACTAC GTCAAGGAAA CAGTAGTGGT AGATGGTCAG 6720

TTGACAACCA GTCGGGGTCC TTCAACAGCC CTTGCCTTTG CCTACGAGTT GGTGGAGCAA 6780

CTAGGAGGGG ACGCAGAGAG TTTACGAACA GGAATGCTCT ATCGAGATGT CTTTGGTAAA 6840

AATCAGTAAA ACGGGAGTTA TTCTCTCGTT TTTTATGTGG AAAACTCAGG GAAATCATCG 6900

CTTTTTTCAT AAAAAAATGC TATAATGAAG GGTATGAAAT ATCACGATTA CATCTGGGAT 6960

TTAGGTGGAA CTTTACTGGA TAATTATGAA ACTTCAACAG CTGCATTTGT TGAAACATTG 7020

GCACTGTATG GTATCACACA AGACCATGAC AGTGTCTATC AAGCTTTAAA GGTTTCTACT 7080

CCTTTTGCGA TTGAGACATT CGCTCCCAAT TTAGAGAATT TTTTAGAAAA GTACAAGGAA 7140

AATGAAGCCA GAGAGCTTGA ACACCCGATT TTATTTGAAG GAGTTTCTGA CCTATTGGAA 7200

GACATTTCAA ATCAAGGTGG CCGTCATTTT TTGGTCTCTC ATCGAAATGA TCAGGTTTTG 7260

GAAATTTTAG AAAAAACCTC TATAGCAGCT TATTTTACAG AAGTGGTGAC TTCTAGCTCA 7320

GGCTTTAAGA GAAAGCCAAA TCCCGAATCC ATGCTTTATT TAAGAGAAAA GTATCAGATT 7380

AGCTCTGGTC TTGTCATTGG TGATCGGCCG ATTGATATCG AAGCAGGTCA AGCTGCAGGA 7440

CTTGATACCC ACTTGTTTAC CAGTATCGTG AATTTAAGAC AAGTATTAGA CATATAAGAA 7500

AAAGGAATAA GATGACAGAA GAAATCAAAA ATCTGCAGGC ACAGGATTAT GATGCCAGTC 7560

AAATTCAAGT TTTAGAGGGC TTAGAGGCTG TTCGTATGCG TCCAGGGATG TACATTGGAT 7620

CAACCTCAAA AGAAGGTCTT CACCATCTAG TCTGGGAAAT TGTTGATAAC TCAATTGACG 7680

AGGCCTTGGC AGGATTTGCC AGCCATATTC AAGTTTTTAT TGAGCCAGAT GATTCGATTA 7740

CTGTTGTGGA TGATGGGCGT GGTATCCCAG TCGATATTCA GGAAAAAACA GGCCGTCCTG 7800

CTGTTGAGAC CGTCTTTACA GTCCTTCACG CTGGAGGAAA GTTCGGCGGT GGTGGATACA 7860

AGGTTTCAGG TGGTCTTCAC GGGGTGGGGT CGTCAGTAGT TAATGCCCTT TCCACTCAAT 7920

TAGACGTTCA TGTTCACAAA AATGGTAAGA TTCATTACCA AGAATACCGT CGTGGTCATG 7980

TTGTCGCAGA TCTTGAAATA GTTGGAGATA CGGATAAAAC AGGAACAACT GTTCACTTCA 8040

CACCGGACCC AAAAATCTTC ACTGAAACAA CAATCTTTGA TTTTGATAAA TTAAATAAAC 8100

GGATTCAAGA GTTGGCCTTT CTAAATCGCG GTCTTCAAAT TTCAATTACA GATAAGCGCC 8160

AAGGTTTGGA ACAAACCAAG CATTATCATT ATGAAGGTGG GATTGCTAGT TACGTTGAAT 8220

ATATCAACGA GAACAAGGAT GTAATCTTTG ATACACCAAT CTATACAGAC GGTGAGATGG 8280

ATGATATCAC AGTTGAGGTA GCCATGCAGT ACACAACTGG TTACCATGAA AATGTCATGA 8340

GTTTCGCCAA TAATATTCAT ACCCATGAAG GTGGAACACA TGAACAAGGT TTCCGTACAG 8400

CCTTGACACG TGTTATCAAC GATTATGCTC GTAAAAATAA GTTACTGAAA GACAATGAAG 8460

ATAATTTAAC AGGGGAAGAT GTTCGCGAAG GCTTAACTGC AGTTATCTCA GTTAAACACC 8520

CAAATCCACA GTTTGAAGGA CAAACCAAGA CCAAATTGGG AAATAGCGAA GTGGTCAAGA 8580

TTACCAATCG CCTCTTCAGT GAAGCTTTCT CCGATTTCCT CATGGAAAAT CCACAGATTG 8640

CCAAACGTAT CGTAGAAAAA GGAATTTTGG CTGCCAAGGC TCGTGTGGCT GCCAAGCGTG 8700

CGCGTGAAGT CACACGTAAA AAATCTGGTT TGGAAATTTC CAACCTTCCA GGGAAACTAG 8760

CAGACTGTTC TTCTAATAAC CCTGCTGAAA CAGAACTCTT CATCGTCGAA GGAGACTCAG 8820

CTGGTGGATC AGCCAAATCT GGTCGTAACC GTGAGTTTCA GGCTATCCTT CCAATTCGCG 8880

GTAAGATTTT GAACGTTGAA AAAGCAAGTA TGGATAAGAT TCTAGCCAAC GAAGAAATTC 8940

GTAGTCTTTT CACAGCCATG GGAACAGGAT TTGGCGCAGA ATTTGATGTT TCGAAAGCCC 9000

GTTACCAAAA ACTCGTTTTG ATGACCGATG CCGATGTCGA TGGAGCCCAC ATTCGTACCC 9060

TTCTTTTAAC CTTGATTTAT CGTTATATGA AACCAATCCT AGAAGCTGGT TATGTTTATA 9120

TTGCCCAACC ACCAATCTAT GGTGTCAAGG TTGGAAGCGA GATTAAAGAA TATATCCAGC 9180

CGGGTGCAGA TCAAGAAATC AAACTCCAAG AAGCTTTAGC CCGTTATAGT GAAGGTCGTA 9240

CCAAACCGAC TATTCAGCGT TATAAGGGGC TAGGTGAAAT GGACGATCAT CAGCTGTGGG 9300

AAACAACCAT GGATCCCGAA CATCGCTTGA TGGCTAGAGT TTCTGTAGAT GATGTGCAGA 9360

AGCAGATAAA ATCTTTGATA TGTTGATGGG GATCGAGTTG TCCTCGTCG 9409

(2) INFORMATION FOR SEQ ID NO: 162:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6415 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

CCTGGGAAAG TCTTGAAAAT TATGATAGAA TGGTGGAAGG AAAAATTCAG GAGAGTAGTA 60

GTGACTCAAA ATGTTGAAAG TCTTCTCGTA TCCATTGTAA TCAGTGCATA CAATGAAGAA 120

AAATATCTGC CTGGTCTAAT TGAAGACTTA AAAAATCAAA CCTATCCTAA AGAGGATATT 180

GAAATTCTAT TTATAAATGC TATGTCCACA GATGGGACCA CAGCTATCAT TCAGCAATTT 240

ATAAAGGAAG ATACAGAGTT TAACTCAATT AGATTGTATA ACAATCCTAA GAAAAATCAA 300

GCTAGTGGTT TTAACCTGGG AGTTAAACAT TCTGTAGGGG ACCTTATTTT AAAAATTGAT 360

GCTCATTCAA AAGTTACTGA GACTTTTGTA ATGAACAATG TGGCTATTAT TCAACAAGGT 420

GAATTTGTCT GTGGGGGGCC TAGACCGACG ATTGTCGAAG GAAAAGGAAA ATGGGCAGAG 480

ACCTTGCATC TTGTTGAGGA AAATATGTTT GGCAGTAGCA TTGCCAATTA TCGAAATAGT 540

TCTGAGGATA GATATGTTTC TTCTATTTTT CATGGAATGT ATAAACGAGA GGTTTTCCAG 600

AAGGTTGGTT TAGTAAATGA GCAACTTGGC CGAACTGAAG ATAATGATAT TCATTATAGA 660

ATTCGAGAAT ATGGTTATAA AATCCGCTAT AGCCCAAGTA TTCTATCTTA TCAGTATATT 720

CGACCAACAT TCAAGAAAAT GCTGCATCAA AAGTATTCAA ATGGTTTGTG GATTGGCTTG 780

ACAAGTCATG TTCAGTTTAA GTGTTTATCA TTATTTCACT ATGTTCCTTG TTTATTTGTT 840

TTGAGTCTTG TGTTTAGTCT AGCATTGTTA CCGATCACAT TCGTATTCAT AACTTTACTA 900

TTAGGTGCCT ATTTTCTACT TTTGTCATTA CTCACTTTGC TGACTTTATT AAAACATAAA 960

AATGGATTTC TAATTGTGAT GCCCTTTATT TTATTTTCCA TTCACTTTGC TTATGGCCTT 1020

GGGACGATTG TAGGTTTAAT TAGAGGATTT AAATGGAAGA AGGAGTACAA GAGAACAATA 1080

ATTTATTTGG ATAAAATAAG CCAAATAAAT CAAAATATGC TATAATAACA ATATAGTAAA 1140

ACTCTTTTAA GGAGGAGTAG ATTTCTATGA ATAAAAAACT AACAGATTAT GTGATTGATC 1200

TGGTGGAAAT TTTAAATAAA CAACAAAAGC AGGTTTTCTG GGGAATATTT GATATTTTCA 1260

GTATGGTGGT TTCCATCATT GTATCTTATA TTTTATTTTA TGGGCTGATT AATCCAGCAC 1320

CTGTTGACTA CATTATCTAT ACGAGTTTGG CCTTCCTGTT CTATCAATTG ATGATTGGTT 1380

TTTGGGGGTT GAACGCGAGC ATTAGTCGTT ACAGCAAGAT TACGGATTTC ATGAAAATCT 1440

TTTTTGGTGT GACTGCTAGC AGTGTCTTGT CATATAGTAT CTGTTATGCC TTCTTGCCAC 1500

TCTTCTCCAT CCGTTTCATC ATTCTCTTTA TCTTGTTGAG TACCTTCTTG ATTTTATTGC 1560

1050

CACGGATTAC TTGGCAGTTA ATCTACTCCA GACGCAAAAA AGGTAGTGGT GATGGAGAAC 1620

ACCGTCGGAC CTTCTTGATT GGTGCCGGTG ATGGTGGGGC TCTTTTTATG GATAGTTACC 1680

AACATCCAAC CAGTGAATTA GAACTGGTCG GTATTTTGGA TAAGGATTCT AAGAAAAAGG 1740

GTCAAAAACT TGGTGGTATT CCTGTTTTGG GCTCTTATGA CAATCTGCCT GAATTAGCCA 1800

AACGCCATCA AATCGAGCGT GTCATCGTTG CGATTCCGTC GCTGGATCCG TCAGAATATG 1860

AGCGTATCTT GCAGATGTGT AATAAGCTGG GTGTCAAATG TTACAAGATG CCTAAGGTTG 1920

AAACTGTTGT TCAGGGCCTT CACCAAGCAG GTACTGGCTT CCAAAAAATT GATATTACGG 1980

ACCTTTTGGG TCGTCAGGAA ATCCGTCTTG ACGAATCGCG TCTGGGTGCA GAACTGACAG 2040

GTAAGACCAT CTTAGTCACA GGAGCTGGAG GTTCAATCGG TTCTGAAATC TGTCGTCAAG 2100

TTAGTCGCTT CAATCCTGAA CGCATTGTCT TGCTCGGTCA TGGGGAAAAC TCAATCTACC 2160

TTGTTTATCA TGAATTGATT CGTAAGTTCC AAGGGATTGA TTATGTACCT GTGATTGCGG 2220

ACATTCAAGA CTATGATCGT TTGTTGCAAG TCTTTGAGCA GTACAAACCT GCTATTGTTT 2280

ATCATGCGGC AGCCCACAAG CATGTTCCTA TGATGGAGCG CAATCCAAAA GAAGCCTTCA 2340

AAAACAATAT CCGTGGAACT TACAATGTTG CTAAGGCTGT TGATGAAGCT AAAGTGTCTA 2400

AGATGGTTAT GATTTCGACA GATAAGGCAG TCAATCCACC AAATGTTATG GGAGCAACCA 2460

AGCGCGTGGC GGAGTTGATT GTCACTGGCT TTAACCAACG TAGCCAATCA ACCTACTGTG 2520

CAGTTCGTTT TGGGAATGTT CTTGGTAGCC GTGGTAGTGT CATTCCAGTC TTTGAACGTC 2580

AGATTGCTGA AGGTGGGCCT GTAACGGTGA CAGACTTCCG TATGACCCGT TACTTTATGA 2640

CCATTCCAGA AGCTAGCCGT CTGGTTATCC ATGCTGGTGC TTATGCCAAA GATGGGGAAG 2700

TCTTTATCCT TGATATGGGC AAACCAGTCA AGATTTATGA CTTGGCCAAG AAGATGGTGC 2760

TTCTAAGTGG CCACACTGAA AGTGAAATTC CAATCGTTGA AGTTGGAATC CGCCCAGGTG 2820

AAAAACTCTA CGAAGAACTC TTGGTATCAA CCGAACTCGT TGATAATCAA GTTATGGATA 2880

AGATTTTCGT TGGTAAGGTT AATGTCATGC CTTTAGAATC CATCAATCAA AAGATTGGAG 2940

AGTTCCGCAC TCTCAGTGGA GATGAGTTGA AGCAAGCTAT TATCGCCTTT GCTAATCAAA 3000

CAACCCACAT TGAATAAAAA AGAAAAACGC ATAGTATCAA GTTACACAAC CTTGGTAATA 3060

TGCGTTTTAT TATGTAGAGA CTTATACTCT TCGAAAATCT CTTCAAACCA CGTCAACGTC 3120

GCCTTGCCGT ATATGGTTAC TGACTtCGTC AGTTCTATCC ACAACCTCAA AACAGTGTTT 3180

TGAGytGACT TCGTCAGTTC TATCCACAAC CTCAAAACAG TGTTTTGAGc TGACttCGTC 3240

AGTTCTATCC ACAACCTCAA AACAGTGTTT TGAGCTGAcT TCGTCAGTTC CATCCACAAC 3300

CTTAAAACAG TGTTTTGAGy TGACnTTCGT CAGTTCCATC TACAACCTTA AAACAGTGTT 3360

TTGAGCTGCC CGCAGCTAGT TTCCTAGTTT GCTCTTTGAT TTTCATTGAG TATTACTTCA 3420

TTTTCTTCTG AAATGGAATT GTTACCCAGT CTATGCTATT GAAAATACGC CAAAACTTCT 3480

AAGGGTTTGT GAGCGATATA ATCAGGTTGA TAGTTTAGTA GATCTGCTTG CTCTCCAAAT 3540

CCCCAAGTGA TGGCCAATTT CTGAATACCT GTTTCTCGAG CTCCCAGCAT ATCAAACTTG 3600

GTATCTCCGA TGATGATGGC TTGTTCTGGT GCTAGTTGAT GTGTCTGCAA GGCTTGGTGA 3660

ATGACATCTG CCTTATGGGG TGCTTCAGGG CTAGAACCAT AAATGCCATC AAAGAAATGA 3720

TGGATTTCCA AGTTTTTTGC CATGTCTTGA GCAGTAGATG TATCCTTTGT CGTGGTGATG 3780

TAGAGTGGAT AACTGCTCGA TAACTCCTCA AGCAAGTCTA TAATCTGAGG AAAGAGTTGA 3840

GCTTCATAGA TGCCTTTTGC CTTATAGTAA GAACGATATA TCTGCACGGC TTCAGAAATT 3900

TGGTCTTTGG ACAGGCAGGT CGCAAAACTA CTTTCGAGAG GTGGTCCCAT AAAACCACGA 3960

ATAGTTTTGG CATCAGGGCT AGGCACCCCC AGCTCTTTAA AGGTATAGGT AAAGGCATTG 4020

TGAATCCCGA TAGAACTATC AACGAGGGTT CCATCCAAAT CGAAAAAAAT CGCTGTGATA 4080

GAGGTCATGG TTTCTCCTAT TTGATAAGCT TATTCTCCGA AAATTTCTTT TTGGAGGCGA 4140

CGACCAGTAG GGGTGGTAGC GAGTCCACCT TCAGCTGTTT CACGAAAGGC AGTTGGCATG 4200

CTTGCTCCTA CTTGGTACAT GGCATCGATC ACTTCATCCA CAGGGATTTT AGATTCGATA 4260

CCTGCCAAGG CCATGTCTGC TGCGATGAAA GCAAAGCTAG CTCCCATGGC ATTACGTTTG 4320

ACACAGGGAA CTTCGACCAA ACCTGCAACA GGGTCACAGA TGAGGCCTAG CATATTTTTA 4380

ATGACAAAGG CAATAGCTTG ACTGGCCTGA TAAGGTGTTC CACCTGCAGC CAGAGTCAAG 4440

GCGGCAGCAC TCATAGCAGA GGCTGAACCA ACTTCAGCTT GACACCCACC CTCAGCACCT 4500

GAGATGGAGG CATTGTTTGC GATGACTAGT CCAAAGGCAC CAGCAGCAAA GAGGAAATCC 4560

AATTGTTGCT CGTGGCTGAG GTCTAATTTT TCAATAGCAG CAGTGAGAAC GGATGGCAGA 4620

CAGCCAGCAC TTCCAGCGGT TGGAGTGGCA CAGACCAAGC CCATTTTGGC ATTGTGTTCA 4680

TTGACTGCGA TGGCATTTCG GGCAGCCGAG AGAATCGTAT AATCTGACAG AGTTTTTCCG 4740

TTTTCGATGT AGTGATCCAA TTTGGCAGCA TCTCCACCTG TCAGGCCACT ACGAGATTTA 4800

TTTTCATTGA GGCCAAGTTG GACAGAGGCT TTCATAACTT CCAGATTGCG TTCCATGAGA 4860

AGGAAGACTT CTTCACGTTC GCGACCGGTC AATTCAAACT CTGTTGTAAT CATGAGTTCT 4920

GCGACATTTC CTTGAAAGTC CAGATCTGCT TGCTCGACCA ATTCTTTGAT AGAATAAAAC 4980

ATGCTTCCTC CTATTTAAAG AAATTGACAT TGTGGAGATG AGGGATTTTT CGAATTTCTT 5040

CGATAGCCTC ATCACAGTTG CGACTGTCAA CTTCGATAAT CATAATGGCT TTTTCACCAG 5100

CTTTTTCACG AGTGACATTC ATCTGGGCGA TATTGATACC ATAGCGGGAA AGCGCCTCTG 5160

TAACAAGGGC AATCATACCT GGAATATCTT GATGAACGAT GATGATAGTC GGTGTATTCA 5220

TATTGAGAGA GACGGCAAAA CCATTGAGTT CGGTTACCTG AATATTTCCT CCACCGATAG 5280

AAATACCAGT CACGCTGATG GTCTTGTGGG CATTTTTAAC AGTAATTTTA GTGGTGTTAG 5340

GGTGAGGGGC ATTGCTGTCT TTCTGAATGG TCCAGACAAT CTTGATACCA CGCTTGTGGG 5400

CAATTTCCAG ACTATTTGGA ATTTCAGGAT CATCTGTATC CATTCCTAAA ATACCTGCAA 5460

CAAGGGCTAG GTCTGTTCCG TGACCACGAT AGGTCTTGGC AAATGAGTTA AAAAGTTGGA 5520

ATTCAACTTC TGTCGGAGTA TCATCAAAAA TGGAAGAGAC AATCTTCCCA ATACGAACAG 5580

CACCAGCGGT ATGGCTACTA GATGGGCCAA TCATAACTGG TCCGATGATA TCAAAGACAG 5640

ATTGAAAACG AAGTGATTTC ATCAGTTTCC CCTTATAAAA ATTCTTATCT CTATTATATC 5700

AAAGAATGAG GGGCTTGGCT TTAATTGTGG ATGAAAACCT TTCTAATACC TCAAATAGCA 5760

TAAAAATAGT ATCTTTTATG ACAAAAAACA CCTTATTTAG GGAAATAAAA AATAATTTTG 5820

TAATATTTCT ACATAAAAGT GTCAAGAAAC GGTAATATTT AAAGGGTATG ATAGAACTAT 5880

AGAAAGAAGG AGAATTTTCG AATATGAAAT CAATAACTAA AAAGATTAAA GCAACTCTTG 5940

CAGGAGTAGC TGCCTTGTTT GCAGTATTTG CTCCATCATT TGTATCTGCT CAAGAATCAT 6000

CAACTTACAC TGTTAAAGAA GGTGATACAC TTTCAGAAAT CGCTGAAACT CACAACACAA 6060

CAGTTGAAAA ATTGGCAGAA AACAACCACA TTGATAACAT TCATTTGATT TATGTTGATC 6120

AAGAGTTGGT TATCGATGGC CCTGTAGCGC CTGTTGCAAC ACCAGCGCCA GCTACTTATG 6180

CGGCACCAGC CGCTCAAGAT GAAACTGTTT CAGCTCCAGT AGCAGAAACT CCAGTAGTAA 6240

GTGAAACAGT TGTTTCAACT GTAAGCGGAT CTGAAGCAGA AGCCAAAGAA TGGATCGCTC 6300

AAAAAGAATC AGGTGGTAGT ATACAGCTAC AAATGGACGT TATATCGGAC GTTACCAATT 6360

AACAGATTCA TACCTGAACG GTGACTACTC AGCTGAAAAC CAAGAACGGG TACCG 6415

(2) INFORMATION FOR SEQ ID NO: 163:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8494 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

TACCCCTTTC GAATTTTGGC AAAAATTCGG TAAGGCTTTG ATGGTAGTTA TCGCGGTTAT 60

GCCGGCTGCT GGTTTGATGA TTTCAATCGG TAAGTCTATC GTGATGATTA ACCCAACCTT 120

TGCACCACTT GTCATCACAG GTGGAATTCT TGAGCAAATC GGTTGGGGGG TTATCGGTAA 180

CCTTCACATT TTGTTTGCCC TAGCCATTGG AGGAAGCTGG GCTAAAGAAC GTGCTGGTGG 240

TGCTTTCGCC GCTGGTCTTG CCTTCATCTT GATTAACCGT ATCACTGGTA CAATCTTTGG 300

TGTATCAGGC GATATGTTGA AAAATCCAGA TGCTATGGTA ACTACTTTCT TTGGTGGTTC 360

AATCAAAGTT GCTGATTACT TTATCAGTGT TCTTGAAGCT CCAGCCTTGA ACATGGGGGT 420

ATTCGTAGGG ATTATCTCAG GTTTTGTAGG GGCAACTGCT TACAACAAAT ACTACAACTT 480

CCGTAAACTT CCTGATGCAC TTTCATTCTT CAACGGGAAA CGTTTCGTAC CATTTGTAGT 540

TATTCTTCGT TCAGCAATCG CTGCAATTCT ACTTGCTGCT TTCTGGCCAG TAGTTCAAAC 600

AGGTATCAAT AACTTCGGTA TCTGGATTGC CAACTCACAA GAAACTGCTC CAATTCTTGC 660

ACCATTCTTG TATGGTACTT TGGAACGTTT GCTCTTGCCA TTTGGTCTTC ACCACATGTT 720

GACTATCCCA ATGAACTACA CAGCTCTTGG TGGTACTTAT GACATTTTAA CTGGTGCAGC 780

TAAAGGTACT CAAGTATTCG GTCAAGACCC ACTATGGCTT GCATGGGTAA CAGACCTTGT 840

AAACCTTAAA GGTACTGATG CTAGTCAATA TCAACACTTG TTAGATACAG TACATCCAGC 900

TCGTTTCAAA GTTGGACAAA TGATCGGTTC ATTCGGTATC TTGATGGGTG TGATTGTTGC 960

TATCTACCGT AATGTTGATG CTGACAAGAA ACATAAATAC AAAGGTATGA TGATTGCAAC 1020

AGCTCTTGCA ACATTCTTGA CAGGGGTTAC TGAACCAATC GAATACATGT TCATGTTCAT 1080

CGCAACACCT ATGTATCTTG TTTACTCACT TGTTCAAGGT GCTGCCTTCG CTATGGCTGA 1140

CGTCGTAAAC CTACGTATGC ACTCATTCGG TTCAATCGAG TTCTTGACTC GTACACCTAT 1200

TGCAATCAGT GCTGGTATTG GTATGGATAT CGTTAACTTC GTTTGGGTAA CTGTTCTCTT 1260

TGCTGTAATC ATGTACTTTA TCGCAAACTT CATGATTCAA AAATTCAACT ACGCAACTCC 1320

AGGGCGCAAC GGAAACTACG AAACTGCTGA AGGTTCAGAA GAAACCAGCA GCGAAGTGAA 1380

AGTTGCAGCA GGCTCTCAAG CTGTAAACAT TATCAACCTT CTTGGTGGAC GTGTAAACAT 1440

CGTTGATGTT GATGCATGTA TGACTCGTCT TCGTGTAACT GTTAAAGATG CAGATAAAGT 1500

AGGAAATGCA GAGCAATGGA AAGCAGAAGG AGCTATGGGT CTTGTCATGA AAGGACAAGG 1560

GGTTCAAGCT ATCTACGGTC CAAAAGCTGA CATTTTGAAA TCTGATATCC AAGATATCCT 1620

TGATTCAGGT GAAATCATTC CTGAAACTCT TCCAAGCCAA ATGACTGAAG CACAACAAAA 1680

CACTGTTCAc TTCAAAGATC TTACTGAGGA AGTTTACTCA GTAGCAGACG GTCAAGTTGT 1740

TGCTTTGGAA CAAGTAAAGG ATCCAGTATT TGCTCAAAAA ATGATGGGTG ATGGATTTGC 1800

AGTAGAACCT GCAAATGGAA ACATTGTATC TCCAGTTTCA GGTACTGTGT CAAGCATCTT 1860

CCCAACAAAA CATGCTTTTG GTATTGTGAC GGAAGCAGGT CTTGAAGTAT TGGTTCACAT 1920

TGGTTTGGAC ACAGTAAGTC TTGAAGGTAA ACCATTTACA GTTCATGTTG CTGAAGGACA 1980

AAAAGTTGCA GCAGGAGATC TCCTTGTCAC AGCTGACTTG GATGCTATCC GTGCAGCAGG 2040

ACGTGAAACT TCAACAGTAG TTGTCTTCAC AAATGGTGAT GCAATTAAAT CAGTTAAGTT 2100

AGAAAAAACA GGTTCTCTTG CAGCTAAAAC AGCAGTTGCT AAAGTAGAAT TGTAATATAC 2160

TTGAGGTTGG AAGCTGTATT CCAACCTCTT ATTTTGGGAG AAAAGAATGA AATTTTTAAC 2220

ACTCAATACT CACAGTTGGA TGGAGAAAGA AGCAGAGGAA AAATTCCAGA TTTTGCTTGA 2280

AGATATTCTT GAAAAGGACT ATGATTTGAT TTGTTTTCAA GAAATCAATC AGGAGATGAC 2340

CTCGTCAGAG GTGGAGGTTA ATGACCTTTA TCAAGCTTTG CCAGCAGCTG AGCCTATTCA 2400

CCAAGACCAT TATGTTAGAC TCTTGGTTGA AAAGTTGTCT GAGCAAGGGA AAAATTACTA 2460

CTGGACCTGG GCCTATAACC ATATCGGCTA TAACCGCTAC CACGAAGGTG TGGCTATCTT 2520

GTCTAAAACA CCTATTGAAG CCAGAGAAAT TTTGGTTTCA GATGTGGATG ATCCAACAGA 2580

CTATCATACT CGCCGTGTTG CCCTAGCTGA AACTGTAGTC GATGGCAAGG AGCTAGCAGT 2640

TGCCAGTGTT CATCTCTCTT GGTGGGATAA AGGTTTCCAA GAAGAATGGG CACGATTTGA 2700

GGCTGTCTTG AAAAAATTGA ACAAGCCACT TTTACTAGCT GGAGATTTCA ACAATCCGGC 2760

TGGACAGGAA GGTTACCAAG CTATTTTAGC TAGTCCATTA GGCTTACAAG ACGCATTTGA 2820

AGTTGCTCAA GAGAAAAGTG GTAGCTATAC TGTTCCGCCT GAAATTGATG GCTGGAAAGG 2880

GAACACTGAA CCCCTTCGAA TCGATTATGT CTTTACTACC AAAGAGTTAG CGGTGGAAAA 2940

TTTACATGTC GTATTTGATG GTAACAAGAG TCCACAAGTG AGTGATCACT ATGGCTTGAA 3000

TGCTATATTA AACTGGAAAT AATAACTGAA AAGAGGTTGG AACTATAAAA TTCCAGCCTT 3060

TTCTTACTAG AGAAGCTACT GGAAATAGCC TAAATAAGTG AGACTACTGT AATGGAATAA 3120

AATATGGTAT AATTGATAAG GTAGATAGAA TCGAGGATGT TATGTCATTT ACGAAATTTC 3180

AATTTAAAAA CTATATTAGA GAAGCCTTGA AGGAGTTAAA ATTTACAACT CCAACAGAGG 3240

TGCAAGACAA GTTGATTCCT ATTGTTTTGG CAGGTCGTGA CCTAGTAGGA GAATCAAAAA 3300

CAGGTTCAGG TAAGACTCAT ACTTTCTTGT TACCGATTTT CCAGCAATTA GATGAAGCTA 3360

GCGATAGTGT ACAAGCAGTG ATTACTGCAC CGAGTCGTGA GTTGGCTACT CAAATTTACC 3420

AAGTAGCGCG TCAGATTTCA GCTCACTCAG ATGTCGAAGT TCGTGTGGTT AATTATGTGG 3480

GTGGTACGGA TAAGGCTCGC CAGATTGAGA AATTGGCAAG CAATCAGCCT CATATTGTTA 3540

TTGGAACACC AGGCCGTATC TACGACTTGG TTAAATCTGG TGATTTAGCT ATTCATAAAG 3600

CCAAGACATT TGTTGTTGAT GAAGCAGATA TGACCTTGGA TATGGGATTC TTGGAAACTG 3660

TTGATAAGAT TGCTGGCAGT CTTCCAAAAG ACTTGCAATT CATGGTCTTC TCAGCGACTA 3720

TCCCACAAAA ACTGCAACCA TTCTTGAAAA AATACTTATC AAATCCTGTT ATGGAGAAAA 3780

TTAAGACCAA AACGGTTATT TCTGACACCA TTGATAATTG GTTGATTTCG ACCAAGGGAC 3840

ATGATAAGAA TGCTCAAATT TACCAGTTGA CTCAGTTGAT GCAGCCGTAT TTGGCAATGA 3900

TTTTTGTTAA CACTAAAACG CGTGCTGATG AATTGCATTC ATATCTGACT GCTCAAGGCT 3960

TGAAGGTTGC AAAAATCCAT GGCGATATTG CCCCTCGTGA ACGCAAGCGA ATCATGAATC 4020

AGGTGCAAAA TCTGGATTTT GAGTATATTG TCGCAACAGA TTTGGCAGCG CGTGGGATTG 4080

ACATTGAAGG TGTCAGCCAT GTCATCAATG ATGCCATTCC GCAAGACTTA TCTTTTTTTG 4140

TTCATCGTGT TGGTCGTACT GGACGAAATG GCCTACCAGG TACAGCTATT ACCCTTTATC 4200

AGCCAAGTGA TGACTCGGAT ATCCGTGAGT TGGAGAAATT GGGAATCAAG TTTAGTCCTA 4260

AGATGGTCAA AGACGGGGAA TTTCAAGATA CCTATGACCG TGATCGTCGT GCCAACCGTG 4320

AGAAAAAACA AGATAAACTT GATATCGAAA TGATTGGTTT GGTTAAAAAG AAAAAGAAAA 4380

AAGTCAAACC GGGTTATAAG AAGAAAATTC AATGGGCGGT TGATGAAAAG CGCCGTAAAA 4440

CCAAGCGTGC TGAAAATCGC GCTCGCGGTC GTGCAGAGCG TAAAGCTAAA CGCCAAACAT 4500

TTTAATAGAA ATTGTTGGAG TATTGAGCTC CAACTTTTTT ATTTATGAGA ACGAACTATC 4560

TAAACCGAAA CACTACATTA AAGACTGCAA ATTGCGATTA AAAATGGTAT AATGATAAAG 4620

TTATATAGTC CCGATAAGAT GGTAGGTATT TATTACGAAG AGTTTTCCTA TCAGTACTTT 4680

GTAACTCTAT AACAATATTT TTTAAGGGGG GACATTTTTA TGTCAGAGCG TAAATTATTC 4740

ACGTCTGAAT CTGTATCTGA GGGGCATCCG GATAAGATTG CAGACCAAAT TTCAGATGCG 4800

ATTTTGGATG CTATTTTAGC AAAGGATCCA GAGGCGCACG TTGCTGCTGA AACAGCTGTA 4860

TATACTGGTT CTGTCCACGT TTTTGGTGAA ATTTCTACAA ATGCCTATGT GGATATTAAC 4920

CGTGTGGTTC GTGATACCAT TGCAGAGATT GGTTATACCA ATACAGAATA TGGATTTTCT 4980

GCTGAGACGG TGGGAGTACA CCCATCTTTG GTGGAACAAT CTCCTGACAT CGCTCAAGGT 5040

GTTAACGAAG CCTTGGAGGT TCGTGGAAAT GCTGATCAAG ATCCACTGGA CTTGATTGGA 5100

GCAGGTGACC AAGGGCTCAT GTTTGGATTT GCAGTAGATG AAACAGAAGA GCTTATGCCA 5160

TTGCCAATTG CACTCAGTCA TAAATTGGTT CGTCGTCTGG CAGAACTTCG TAAGTCTGGA 5220

GAAATTAGCT ATCTCCGTCC AGATGCAAAA TCACAAGTTA CAGTTGAGTA CGATGAAAAT 5280

GACCGTCCGG TACGTGTAGA TACAGTCGTT ATTTCTACTC AGCATGATCC AGAGGCCACT 5340

AATGAACAAA TCCATCAAGA TGTGATTGAC AAGGTCATCA AAGAAGTTAT TCCATCTTCT 5400

TATCTTGATG ATAAGACAAA ATTCTTTATC AATCCGACAG GTCGTTTTGT AATCGGTGGT 5460

CCTCAAGGGG ACTCAGGTTT GACTGGTCGT AAGATTATTG TAGATACTTA TGGTGGCTAC 5520

TCTCGTCATG GTGGTGGTGC CTTCTCTGGT AAAGATGCGA CTAAGGTGGA TCGTTCAGCC 5580

TCTTATGCGG CTCGCTATAT TGCCAAGAAT ATCGTTGCAG CAGACCTTGC TAAGAAGGCA 5640

GAAGTGCAGT TGGCCTATGC TATCGGTGTT GCGCAACCTG TTTCTGTTCG TATCGATACT 5700

TTCGGTACAG GAACAGTAGC TGAAAGTCAA CTTGAAAAAG CGGCTCGTCA AATCTTTGAC 5760

CTTCGCCCTG CAGGGATTAT CCAAATGCTG GACCTCAAGC GTCCAATTTA CCGTCAAACA 5820

TCGGCTTACG GTCACATGGG ACGTACAGAT ATTGATCTTC CATGGGAACG TTTGGATAAG 5880

GTAGATGCTT TGAAAGAAGC AGTAAAATAA GATTTTAAGA GGGGAACGTC CTCTCTTTTT 5940

TATAGTTTTT AACTATACTG GGATACTGTT CTGAAAATCC ATTTTGCGAA AGTAGAGATT 6000

TACATGTATA GTAGATTGAA ACTAGAATAG TACACCTCAA CTTCTAAAAC ATTGTTAGCA 6060

ATCAATTTGA CTGTCCTGAT CGATTTCTCC TGTTCTTGTT TCATTTTACT ATATTTCTTT 6120

AAAAATGATA AAGGTTAAGA TTTCTCCTCG TAATAGATAA TCTTGGGGAT ATTTCAATCC 6180

AAAGTTTTAT TCGTTATCAC TTGACTATTG CAAGGTTTTC TAGAGCAACA GAGTCATGGA 6240

ATGGACTCAT GGTTGAGATT TCTCCTTGTT GCTTGGACTT CATTCAAAAG TCTGTTACCC 6300

AAGCCTTGTT CAAACTTCTA ATACACTAGC TGTTTCCATA GCATGACTTC TGTACTAGAC 6360

TTTCTTTTCC GAATAAATAG ATAGAACCAC AGAATCTAGT AAACCTAGAA TTAAAATTAT 6420

GGTATAATAT TAGCAATAAA AGAAATCTGG AGGATTAGAA TCATGGTATC AACGAAAACA 6480

CAAATTGCTG GTTTTGAGTT TGACAATTGC TTGATGAATG CAGCAGGTGT GGCTTGTATG 6540

ACGATAGAGG AGTTAGAAGA GGTCAAAAAC TCAGCGGCAG GAACCTTTGT TACTAAGACA 6600

GCGACCTTGG ACTTCCGTCA GGGGAATCCT GAGCCACGCT ACCAAGATGT TCCACTTGGT 6660

TCCATCAACT CTATGGGCTT GCCAAATAAT GGCTTAGACT ATTATTTGGA TTATCTTTTA 6720

GATTTGCAGG AAAAAGAGTC GAACCGAACT TTCTTCTTAT CTCTGGTCGG CATGTCTCCA 6780

GAGGAAACCC ATACTATTTT GAAAAAAGTC CAAGAGAGTG ATTTTCGTGG TCTGACTGAG 6840

CTAAATCTTT CCTGTCCAAA TGTTCCAGGT AAACCTCAGA TTGCCTATGA TTTTGAGACA 6900

ACAGACCGGA TTTTGGCAGA AGTGTTTGCT TACTTCACCA AACCTCTTGG AATTAAATTG 6960

CCACCTTATT TTGATATTGT TCACTTTGAC CAAGCGGCAG CTATTTTCAA CAAATATCCG 7020

CTCAAGTTTG TCAACTGCGT TAACTCTATC GGAAACGGCC TCTATATAGA AGACGAATCT 7080

GTCGTTATTC GGCCTAAGAA TGGTTTTGGT GGAATTGGTG GAGAATACAT CAAACCGACT 7140

GCTTTAGCCA ATGTTCACGC CTTTTATCAA CGTTTAAATC CTCAAATCCA AATTATCGGA 7200

ACAGGTGGCG TTCTGACTGG TCGAGATGCC TTTGAACACA TCCTCTGTGG AGCAAGTATG 7260

GTGCAGGTGG GAACGACCCT TCACAAAGAA GGCGTCAGTG CTTTTGACCG CATTACCAAT 7320

GAACTGAAAG CAATCATGGT GGAAAAAGGC TACGAGAGCT TAGAAGATTT CCGTGGGAAA 7380

TTGCGCTATA TTGACTAAAT TAAATCGAAA AATCTGAAGA AAGGAGAGAC GATGCTAGCC 7440

ATTGAAGAAA GTCAGAAGTT GACTTTATCA AATTTACCGA GCCTGAGCCT ATTTACAGGG 7500

ACAGATCAGG GTCAGTTTGA AGTGATGAAG AGTCAAATGT TGAAACAGAT TGGGTATGAT 7560

TCTGCTGACC TCAACTTTGC CTACTTTGAT ATGAAAGAAG TAGTTTACAA GGATGTGGAA 7620

CTGGAGTTGG TCAGCCTTCC TTTCTTTGCG GATGAGAAAA TCGTGATATT AGATTATTTT 7680

ATGGATATCA CGACTGCTAA GAAACGCTTT TTGACAGATG ATGAGCTTAA GTCATTTGAG 7740

GAATACCTTG ACAATCCTTC TCCAACAACC AAGTTGATAA TCTTTGCAGA AGGAAAGCTG 7800

GATAGCAAAA GACGGTTAGT CAAATTACTT AAGCGTGATG CCAAGGCCTT CGATGCAGTA 7860

GAAGTAAAAG AACAAGAATT GCGCCAGTAC TTCCAAAAGT GGAGTCAGAA ACAAGGTCTG 7920

CAGTTTACCA ATCATTCTTT TGAAAATCTC CTCATCAAGT CGGGGTTTCA ATTTAGCGAA 7980

ATCCAGAAAA ATCTTCTCTT TTTACAGTCC TATAAGGCGA ATTCTGTTAT TGAGGAAGAG 8040

GATATTGTTA ACGCAATTCC CAAGACTTGC AGGACAATAT TTTTGATTTA ACTCAGTTTA 8100

TTCTGACTAA AAAGATGGAT CAGGCGCGCG ATTTGGTGAG AGACTTGACC TTGCAAGGGG 8160

AAGATGAAAT CAAACTGATT GCAGTCATGC TGGGACAATT TCGGACTTTT ACTCAGGTGA 8220

AGATTTTGGC GGAGTCTGGC CAAACAGAAT CGCAGATTGC AAGTAGTTTA GGTAGTTATC 8280

TGGGACGTAA CCCAAATCCT TATCAAATCA AGTTTGCATT AAGAGATTCG AGAGGACTTT 8340

CTTTGAGCTT TTTGAAGCAA GCTATTTCCT ATTTGATTGA GACAGACTAT CAGATTAAGA 8400

CAGGTCTTTA TGAAAAAGGT TTCCTTTTTG AAAAGGCACT CTTACAGATT GCTAGTCAGG 8460

TCAATTGACA TTTGTTGAAA CTACTAACCC GCGG 8494

(2) INFORMATION FOR SEQ ID NO: 164:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9707 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

CCGGTCAGTT CGTTCAGTAC AAGGAATCAT AATGAACGAT CAATCAGAAA AAAAGACTAG 60

AAAGAAGACT GTATGGATAA TCGACCAATT GGTTTTTTGG ATTCGGGTGT CGGGGGCTTG 120

ACCGTTGTGC GCGAGCTCAT GCGCCAGCTT CCCCATGAAG AAATCGTCTA TATTGGAGAT 180

TCGGCGCGGG CGCCCTATGG CCCCCGTCCT GCTGAGCAAA TTCGTGAATA TACTTGGCAG 240

CTGGTCAACT TTCTCTTGAC CAAGGATGTC AAAATGATTG TCATTGCTTG TAACACTGCG 300

ACTGCGGTCG TCTGGGAAGA AATCAAGGCT CAACTAGATA TTCCTGTCTT GGGTGTAATT 360

TTGCCAGGAG CTTCGGCAGC CATCAAGTCC AGTCAAGGTG GGAAAATCGG AGTGATTGGA 420

ACGCCCATGA CGGTACAATC AGACATATAC CGTCAGAAAA TCCATGATCT GGATCCCGAC 480

TTACAGGTGG AGAGCTTGGC CTGTCCCAAG TTTGCTCCCT TGGTTGAGTC AGGTGCCCTG 540

TCAACCAGTG TTACCAAGAA GGTGGTCTAT GAAACCCTGC GTCCCTTGGT TGGAAAGGTG 600

GATAGCCTGA TTTTGGGCTG TACTCATTAT CCACTCCTTC GCCCTATTAT CCAAAATGTG 660

ATGGGGCCAA AGGTTCAGCT CATCGATAGT GGGGCAGAGT GCGTACGGGA TATCTCAGTC 720

TTACTCAATT ATTTTGAAAT CAATCGTGGT CGCGATGCTG GACCACTCCA TCACCGTTTT 780

TACACAACAG CCAGTAGCCA AAGTTTTGCA CAAATTGGTG AAGAATGGCT GGAAAAAGAG 840

ATTCATGTGG AGCATGTAGA ATTATGACAA ATAAAATTTA TGAATATAAG GATGACCAGG 900

ACTGGTATGT TGGGTCTTAT AGTATTTTTG GTGGCGTTAA CAGTTTGAGC GACTATAAGA 960

CAGATTTTCC TCTGTTTGAA TTCTCCAAAA TATTTGGAGA TGAAGAGTAT GGTTTCCCGC 1020

TTTCAGTTAC TGTTTTACGC TATGGTTCTA TCTACCGTTT GTTCTCCTTT GTGGTAGACA 1080

TGCTTAATCA AGAAATGGGA CGAAACTTGG AAGTTATTCA ACGTCATGGG GCCCTGCTCT 1140

TGGTTGAAAA TGGGCAACTC TTGTATGTAG AATTGCCTAA AGAAGGGGTC AATGTTCATG 1200

ATTTCTTTGA GACAAGCAAG GTCAGAGAAA CCTTGTTGAT TGCGACTCGT AACGAAGGTA 1260

AAACCAAGGA ATTCCGAGCT ATCTTTGATA AGTTAGGCTA CGATGTGGAA AATCTTAATG 1320

ACTACCCTGA CCTGCCTGAA GTAGCAGAAA CAGGTATGAC CTTTGAAGAA AATGCCCGCC 1380

TTAAGGCAGA AACCATTTCT CAATTAACGG GCAAGATGGT TTTGGCAGAT GATTCTGGTC 1440

TCAAAGTCGA TGTCCTTGGT GGCTTACCAG GCGTCTGGTC AGCTCGTTTC GCAGGTGTGG 1500

GAGCAACTGA CCGTGAAAAT AATGCCAAAC TCTTGCACGA ATTGGCCATG GTCTTTGAAC 1560

TCAAGGACCG CTCGGCTCAG TTCCACACAA CCCTAGTCGT AGCCAGCCCA AATAAGGAAA 1620

GTTTAGTTGT TGAAGCAGAC TGGTCAGGTT ATATTAACTT TGAACCTAAG GGTGAAAATG 1680

GCTTTGGCTA TGATCCCCTC TTCCTTGTAG GAGAAACAGG TGAGTCATCA GCTGAATTAA 1740

CCCTGGAAGA AAAAAATAGT CAATCTCACC GTGCCTTAGC CGTTAAGAAA CTTTTGGAGG 1800

TATTTCCATC ATGGCAAAGC AAACCATCAT TGTAATGAGC GATTCCCATG GCGATAGCTT 1860

GATTGTGGAA GAAGTCCGTG ATCGCTATGT GGGCAAAGTC GATGCTGTTT TTCATAACGG 1920

CGATTCTGAA CTACGTCCGG ATTCTCCACT TTGGGAGGGC ATCCGCGTTG TTAAAGGGAA 1980

CATGGACTTC TACGCCGGCT ACCCAGAACG TCTGGTGACT GAGCTTGGTT CGACCAAGAT 2040

TATCCAAACT CATGGTCACT TGTTTGACAT CAATTTCAAC TTTCAAAAGT TGGACTACTG 2100

GGCTCAGGAG GAAGAGGCCG CTATCTGCCT CTATGGTCAC TTGCATGTGC CAAGTGCTTG 2160

GTTGGAAGGC AAGATCCTCT TTCTAAATCC AGGTTCTATC AGTCAACCAC GAGGTACCAT 2220

CAGAGAATGT CTCTATGCTC GTGTGGAGAT TGATGATAGT TACTTCAAAG TGGACTTTTT 2280

GACACGAGAT CACGAGGTGT ATCCAGGTTT GTCCAAGGAG TTTAGCCGAT GATTGCCAAG 2340

GAGTTTGAGA CTTTCTTGTT GGGGCAGGAG GAAACTTTTT TGACCCCTGC TAAAAATCTA 2400

GCTGTGTTGA TTGATACCCA CAATGCGGAT CATGCGACCC TCTTGCTCAG TCAGATGACC 2460

TATACCCGTG TTCCCGTTGT GACAGATGAA AAACAGTTTG TTGGGACGAT TGGACTCAGA 2520

GATATTATGG CTTATCAGAT GGAGCATGAC TTGAGCCAAG AAATCATGGC GGATACGGAT 2580

ATCGTTCATA TGACAAAAAC GGACGTAGCG GTTGTTTCGC CTGATTTCAC CATTACGGAG 2640

GTCTTGCACA AGCTAGTAGA TGAGTCCTTC TTACCGGTTG TGGATGCAGA GGGTATTTTC 2700

CAAGGGATTA TTACGCGCAA GTCCATCCTC AAGGCCGTTA ATGCCCTCTT GCATGACTTT 2760

AGTAAGGAAT ATGAGATTCG ATGCCAATGA GAGACAGGAT TTCAGCCTTT TTAGAGGAAA 2820

AGCAGGGCTT GTCTGTCAAT TCCAAGCAGT CCTATAAGTA TGATTTGGAG CAATTTTTAG 2880

ACATGGTAGG TGAGCGGATT TCTGAGACCA GTCTCAAGAT TTACCAAGCC CAGCTAGCCA 2940

ATCTAAAAAT CAGCGCCCAG AAGCGAAAGA TTTCGGCCTG TAACCAATTT CTATACTTTC 3000

TCTATCAAAA AGGAGAGGTG GACAGCTTTT ACCGCTTGGA ATTAGCCAAA CAAGCTGAAA 3060

AGAAGACGGA AAAGCCAGAG ATTCTATACC TAGACTCTTT TTGGCAGGAA AGCGACCATC 3120

CAGAGGGCCG CTTGCTAGCG CTCTTAATCC TAGAAATGGG GCTCTTGCCC AGTGAGATTT 3180

TAGCCATCAA GGTTGCGGAC ATCAATCTGG ATTTTCAGGT GTTGCGAATC AGCAAGGCTT 3240

CCCAACAGAG GATTGTCACC ATTCCCACGG CCTTGCTTTC AGAATTGGAA CCCTTGATGG 3300

GGCAGACCTA TCTTTTTGAA AGAGGAGAGA AACCCTATTC TCGTCAGTGG GCCTTTCGTC 3360

AGTTAGAATC TTTTGTCAAG GAGAAAGGTT TTCCATCCTT ATCAGCTCAA GTCTTACGTG 3420

AACAGTTTAT TCTAAGACAA ATAGAAAACA AGGTCGATTT GTACGAAATT GCAAAAAAAT 3480

TAGGATTAAA AACAGTCCTG ACCTTAGAAA AATATAGATA ATGGATATTA AATTAAAAGA 3540

TTTTGAAGGA CCCCTGGACT TGCTCTTGCA TCTGGTTTCT AAGTACCAGA TGGATATCTA 3600

1060

CGATGTGCCC ATTACGGAAG TCATCGAACA GTATCTAGCC TATGTCTCAA CCCTGCAGGC 3660

CATGCGTCTG GAAGTGACGG GTGAGTACAT GGTCATGGCT AGTCAGCTCA TGCTGATTAA 3720

GAGTCGTAAA CTCCTTCCGA AGGTAGCAGA AGTGACAGAC TTGGGGGATG ACCTGGAGCA 3780

GGACCTCCTC TCTCAAATCG AAGAATATCG CAAGTTCAAG CTCTTGGGTG AGCACTTGGA 3840

AGCCAAGCAC CAAGAACGGG CCCAGTATTA TTCCAAAGCG CCGACAGAGT TGATTTACGA 3900

AGATGCGGAG CTTGTGCATG ACAAGACGAC CATTGACCTC TTTTTGACTT TTTCAAATAT 3960

CCTAGCCAAG AAAAAAGAGG AGTTTGCACA AAATCACACG ACGATCTTGC GGGATGAGTA 4020

TAAGATTGAG GACATGATGA TTATCGTGAA AGAGTCCTTG ATTGGACGAG ATCAATTGCG 4080

CTTGCAGGAT TTGTTCAAGG AAGCCCAGAA TGTCCAAGAG GTCATCACCC TCTTTTTGGC 4140

AACCCTAGAG TTAATCAAAA CCCAGGAGTT GATCCTCGTG CAAGAGGAGA GTTTTGGAGA 4200

TATCTATCTC ATGGAAAAGA AGGAAGAAAG TCAAGTGCCT CAAAGCTAGA CTTGATAGAG 4260

AGGAAAGATG AGTACTTTAG CAAAAATAGA AGCGCTCTTG TTTGTAGCGG GTGAAGATGG 4320

GATTCGGGTC CGCCAGTTAG CTGAACTCCT CTCTCTGCCA CCGACAGGCA TCCAGCAAAG 4380

TTTAGGAAAA TTAGCCCAGA AGTATGAAAA GGACCCAGAT TCCAGTTTGG CTTTGATTGA 4440

GACAAGTGGT GCTTATAGAT TGGTGACCAA GCCTCAATTT GCAGAGATTT TGAAGGAATA 4500

CTCTAAGGCG CCTATCAACC AGAGCTTGTC TCGGGCTGCC CTTGAGACCT TGTCCATTAT 4560

TGCCTACAAA CAGCCGATTA CGCGGATAGA AATTGATGCC ATCCGTGGAG TTAACTCGAG 4620

TGGAGCCTTG GCAAAGTTGC AGGCTTTTGA CCTGATAAAG GAAGACGGGA AAAAGGAAGT 4680

ATTGGGGCGC CCCAACCTCT ATGTGACTAC GGATTATTTC CTAGATTACA TGGGGATAAA 4740

CCATTTAGAA GAATTACCAG TGATTGATGA GCTTGAGATT CAAGCCCAAG AAAGCCAATT 4800

ATTTGGTGAA AGGATAGAAG AAGATGAGAA TCAATAAGTA TATTGCCCAC GCAGGTGTGG 4860

CCAGTAGGAG AAAAGCAGAA GAGCTGATTA AGCAAGGCTT GGTGACGGTT AACGGCCAAG 4920

TGGTGCGTGA ACTAGCAACC ACTATCAAGT CAGGCGACAA GGTCGAAGTT GAAGGTCAAC 4980

CTATCTACAA CGAAGAAAAG GTCTACTATC TGCTTAACAA ACCACGCGGT GTGATTTCCA 5040

GTGTGACAGA TGATAAGGGT CGCAAGACGG TTGTCGACCT CTTGCCCAAT GTCAAAGAGC 5100

GTATTTACCC TGTGGGTCGT TTGGACTGGG ATACATCAGG TGTCTTGATT TTGACCAATG 5160

ATGGGGACTT TACAGACGAG ATGATTCACC CTCGTAATGA GATTGACAAG GTTTATGTCG 5220

CGCGTGTTAA AGGTGTGGCC AATAAGGACA ATCTCCGCCC CTTGACCCGT GGTCTTGAGA 5280

TTGATGGTAA GAAAACCAAG CCAGCTGTTT ATGAAATTCT CAAAGTGGAC CCAGTCAAAA 5340

ATCGCTCTGT GGTGCAGTTG ACCATCCATG AAGGGCGTAA CCATCAGGTT AAAAAGATGT 5400

TTGAAGCTGT TGGTCTCCAA GTAGATAAGT TGTCTCGGAC TCGTTTCGGA CACCTAGACT 5460

TGACAGGACT CCGTCCAGGA GAATCCCGTC GTCTTAATAA AAAAGAAATC AGCCAACTAC 5520

ACACCATGGC TGTAACTAAG AAATAATGAA ACGAATTTTA ATAGCGCCTG TGCGCTTTTA 5580

CCAACGTTTT ATCTCACCAG TCTTTCCACC CTCTTGTCGC TTTGAGCTGA CTTGTTCCAA 5640

CTACATGATT CAGGCTATTG AAAAACATGG GTTTAAGGGG GTATTGATGG GCTTGGCTCG 5700

GATTTTACGT TGTCATCCCT GGTCGAAAAC AGGTAAGGAC CCCGTTCCAG ACCGCTTTTC 5760

CCTTAAACGA AATCAAGAAG GGGAATGAGG TGGGGTAAAT AGATTTCAAA ATGATAAAAA 5820

CGCATCCTAT CAGGTTTGAG TGAACTTGAT AGGATGCGTT TTAGAATGTC AAAATTTTAT 5880

ACTCTTCGAA AATCTCTTCA AACCGCGTCA GCTTTCATCT GCAACCTCAA AACAGTGTTT 5940

TGAGCAACCT GCGGCTAGTT TCCTAGTTTG CTCTTTGATT TTCATTGAGT ATTAAATTGA 6000

GTTTGAAGTG GCTTATTTCA AAGCTTTTTG TATGTCTTCA ATCATGAGTT TTGTTGATTC 6060

AAGTCCGCCT CCGCTTAGAT ACCAGAGGTC TGGTGTTAGT TGGATAATCT TACCATTTTT 6120

AGCAGCAGGT GTTTCAGCGA TAAGGGCATT TTCTAGGACA CCGTCGTTGC TAGAGTTGTC 6180

CCCACCGATG GCAAGGGTAC GGTTGATGAC AAAGAGGATG TCAGGGTTGA TTTCTTTGAC 6240

ACTTTCAAAG CTGACTTCTT GTCCGTGGCG TGAGTCTTCA AATTTTGTAT CAGTTGGTTT 6300

GAATTTCAAG GTTTGGTACA AGAAAGAGAA ACGAGATTTG GCACCAAAGG CTGCCATTTT 6360

TCCTTCATTA AGGAGGATCG CAAGGGCTTT TTTGTCAGAG CTTTCATTTT TAGTAGCGAC 6420

TTCTTGGATG CTCTTGTCTA GCTTGGTCAA TTCTTCCTTG GCTTTCTGTG TACCAGTTTC 6480

GCCGAAGGCA CTTGCTAAGG ATTCGATATT AGCCTTGGTA GAAGTCCAGT AGTCGTCCTT 6540

GCTTGCTTGG AAGAGAACGG TTGGGGCGAT TTCTTTGAAT TTGTCTACGA ATTTTTGTGT 6600

ACGTGGCGAA GCGATAATCA AATCAGGCTC AAGGGCGGCG ATAGCTTCTA AATCAGGTTC 6660

TTTCATAGAA CCAACATTTT TGACAGTTCC CACTAGGTCT TTTAGATAAG TCGGAACAGT 6720

TTTTGTAGGC ATTCCGACGA TATTTTTTTC AAATCCTAAA GCGCGAATAG TATCCGCAGC 6780

GCCGAGGTCA AAGGTCACAA TCTTTTCAGG AACTTTGGAA AGTTTGACCT CGTCCAGTGA 6840

ACTTTTAATG GTTACCTCTG TTGGAGCAGA GCTACTGGTC TCTGTCTGAC TAGTGCTTGA 6900

GTTTGTACTA CATGCACCAA GTAGGAGCAA GAAGCTGGCC ACTAGGGCAG TGAAATAAAG 6960

TTTAAGGGAT GTTTTCATAA TTTCTCCTTT TTAAAATGTG ATAACGATTT AGGGAGTCTC 7020

TTAATCTTAT TGACTAAGAG ACTGAAGGTT CTCTAACTTG AGCTTTTATG TTACTAGCTA 7080

TAGATACAGA TCTTTTTGTC ATTGATATCA GCTAGCGTGA TGGGAATCTC ATAAAGTTGA 7140

CTGAGCAGGT CAGCCTGCAT GATTTGATCG GTTCTTCCCT TGCTAAAGAC CTGGCCGTCC 7200

TTGAAGGCGA CAATTTCATC TGCATACTGA CTGGCCATGT TGATATCGTG GAGGACGATG 7260

ATAATGGTCT TGCCGAGTTC CTCCACCAGT CGTCGAAGAA TCTGCATCAT GCTGACGCTT 7320

TGCTTGATAT CGAGATTGTT GAGTGGTTCG TCCAGCAAGA TAAAGTCCGT ATCCTGGGCC 7380

AGTACCATAG CGATAAAGAC GCGCTGGAGT TGCCCCCCTG ACAGGCTATT GATGTAGCGG 7440

TCTTTTAAGT TGGTCAGTTC TAAATAGTTC AGAGTTTCTC GGATTTTTTC CCAGTCTTCT 7500

GATCTAAGTC GACCTCGGCT GTAGGGAAAA CGTCCAAAAC TGACCAGTTC TTCAACAGTC 7560

AATTTGGCTT GGTAATTGAT TTTCTGTTTT AGGATGGTTA GTTCTTGGGC CAGTTCTTGC 7620

GAATTCCAGC TCTCGATTTC ACGTCCTTTG ATACTGAGAA CTCCCTGATC TTTCTTGGTT 7680

AGCCTGCTCA TGATGGAGAG GAGAGTCGAT TTTCCAGCAC CATTTGGACC AATAAAGGCT 7740

GTCAGTTTTT GAGGACTGAC TTCAAGCGAA ATGCCTTGCA AAATATCCTG TTTTTGAATG 7800

GATTTGTCAA TGTTTTCCAG TTTCACTGAC GAGACCTCCT ATATAGTAAG ATAAAGAATA 7860

AGAAGCCACC CACACTCTCA ATGATCATAC TGATACGAAT TTCCAGTGCA AAGACTCGTT 7920

CAATCAAGGC TTGCCCCAAG GTTAAGCTAA TAAATCCAAC CAGAATGGCC ACTATAAAGA 7980

GTAACTTGTG CTGATAGTCT TTGACAATCA GGTAGGTGAG GTTGGCCAGT ATAAAGCCGA 8040

AGAAGGCCAT AGGTCCTACC AAGGCAGTGG CCGTTGAGGT CAAAAGCACG ATTCCCCAGA 8100

GGAGCTCTTT CTGTTCTTTT TCAACATCGA GTCCCAATAT CTGAGCCGTT TCTCTTTGCA 8160

GGTGCAAGAC ATCTAGAACG ACTGCTTTTC GAAAGAAAAA GATTGTCAAA GCGAGGATGA 8220

TCAGAGAACC GATGGCTAGG ATGGAAGTGT TGAGATGTTG AAAGGAGGCA AAAAGACTAT 8280

TTTGCAGTTT ATCGTATTCG TTTGGATCCA TTAGGACTTG AAGGAAGGTG CTGATATTTC 8340

GAAAGAGACT TCTGAGCGCT AGACAGATCA GCAGGACGAA GACCAGGTCT TGCTTCATCA 8400

GTGTCTTCAA GTAACCTTGT AAGGCGAGAA AGAAGAGGGA CTGGACAAGA AGTAAGACTA 8460

GGAATTCTAA GATAGGGGAT TTGCCAAGTT GAAGAAACTT GCTTTCAAAA ACCAGTAGTA 8520

GGGTTTGTAG TAGGACGTAG AAGGATTCAA TTCCCAAAAT ACTAGGCGTC AGGAAGCGAT 8580

TTTCCGTCAG GGTTTGAAAA CTAATGGTCG AAATCCCAGT CGCGATGGCT ACCAAGAGAT 8640

AAACGATGAT CTTTTGGGAA CGCAACTTCC AAGCAAAGGC TGACAAGTGA GTGATGGGCC 8700

AAAAGTAGAG AAGACAAGCT CCGATGGCAA GAATAATGAG AATCCAGAAG AGCTTGGTAT 8760

GTTTGCTTTT AGTCTGCATC TTTTCGTCCC CCTCTCCAGA GAAGTAGGAT AAAGACGAGA 8820

CTACCGATGA TTCCTAGCAA GAGACTGACA GACAACTCAT AGGGCCTAAT CAGAACTCGG 8880

GATAGGATAT CGCAAGCCAG AACTAGATTG GCACCAACCA GTGCGACCAT GAGTTTGGTT 8940

TGACTTAGAT TATCTCCATA GCGCTTGCGA ACAAGATTGG GAACGATAAC TCCGAGAAAT 9000

GGTAGGCCAC CCACGGTAAT CATGGTGACG CTTGTCGTTA GCGCCACCAG AAAGAGGGCC 9060

AGTTTTTCAA GTAGGGAGTA GGAAATCCCC AAACTCTCGC TGGTTTCTTT CCCTAGATTC 9120

ATGATGGTGA AGGTTTGGGA TAATTTCCAA ACGGTTATCA GGATGATGAG GCCTAAGAAG 9180

AGCCACTCAT ACTGATGGGT CTGAATCATG GAGAAGGAGC CCTGGGTCCA GGCAGTCATA 9240

CTCTGAACCA GATTGAAACG ATAGGCGATA ACTTCTGTGA CTGAGCCGAT AATCCCGCTA 9300

TAGATGATCC CAATCAGAGG CAACATCCAC CTTTCCTTTA CAGTAAAAAT GGTCATAAAG 9360

GCTAGGAAGA AGAGGGTGAA TACGATGGAT GAAACAAAAG CGAAGAGCAT CTTGTGGGTC 9420

AGACTAGCCG ATGGAAAGAC AAAAAGGCTC AGCACCATTC CCAGTTTGGC GGCTTCAGTC 9480

GTTCCAACTG TACTCGGTGC AGCAAACTGA TTTTGGGTAA TAGTCTGCAT GAGAAGGCCT 9540

GCCATACTCA TACTAGAGGC AGTCAGGAGA ATACTGATAG TTCTTGGGAG ACGGGACTCT 9600

TGAAAGAGGA GCCAGGTCTG CTGGTCGAAA TCAAATAGCT TTCCCCATGA AAAATCACTG 9660

GTCCCAATGC TAATAGAGAG AAAGACTAGG AGTAGAAGTA AGCCAGG 9707

(2) INFORMATION FOR SEQ ID NO: 165:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5910 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

CCGCAATTAT GCTTGAAAAG GAGTATACTT ATAAGTAACG CAAACGTTTG CGTCTGAAAA 60

ATACGCAACG TTCCATTATT TTAACACACG AGGTGCTATT ATGAAAAAAC GTCAAAGTGG 120

TGTGTTGATG CACATCTCTT CTCTTCCAGG AGCTTACGGA ATCGGATCAT TTGGTCAAAG 180

TGCTTACGAC TTCGTTGATT TCTTGGTCCG TACAAAACAA CGTTACTGGC AAATCCTTCC 240

ATTAGGAGCA ACTAGTTACG GGGATTCTCC TTACCAATCT TTCTCAGCCT TCGCAGGAAA 300

CACTCATTTT ATCGATTTAG ATATCTTGGT GGAGCAAGGT TTGTTGGAAG CAAGTGACCT 360

TGAAGGAGTT GACTTTGGTA GCGATGCGTC TGAAGTTGAC TATGCTAAAA TCTACTATGC 420

ACGTCGTCCT CTTTTAGAAA AAGCGGTGAA ACGTTTCTTT GAAGTCGGAG ATGTTAAAGA 480

TTTTGAGAAA TTTGCTCAAG ACAACCAATC ATGGCTTGAG CTCTTTGCTG AGTATATGGC 540

TATCAAAGAG TATTTTGACA ATCTTGCTTG GACTGAATGG CCAGATGCAG ATGCTCGTGC 600

TCGTAAAGCT TCAGCACTTG AAAGCTATCG TGAGCAATTG GCAGACAAGT TGGTTTACCA 660

CCGTGTGACT CAATACTTCT TCTTCCAACA ATGGTTGAAA TTGAAAGCTT ACGCTAACGA 720

CAACCACATC GAAATCGTTG GGGACATGCC AATCTACGTA GCGGAAGATT CAAGTGATAT 780

GTGGGCAAAT CCACATCTCT TCAAAACAGA TGTCAATGGT AAGGCTACTT GTATCGCAGG 840

ATGCCCACCA GATGAGTTTT CTGTAACTGG TCAGCTTTGG GGTAATCCAA TCTATGACTG 900

GGAAGCAATG GACAAAGACG GCTACAAATG GTGGATTGAA CGCTTGCGTG AAAGCTTCAA 960

AATCTACGAT ATCGTTCGTA TCGACCACTT CCGTGGCTTC GAATCTTACT GGGAAATCCC 1020

TGCTGGTTCC GATACAGCAG CACCTGGTGA GTGGGTGAAA GGTCCAGGTT ACAAGCTTTT 1080

TGCAGCCGTT AAGGAAGAAC TTGGTGAGCT AAACATCATC GCAGAAGACC TTGGCTTCAT 1140

GACAGATGAA GTGATCGAAT TGCGTGAACG TACTGGCTTC CCAGGAATGA AGATTCTTCA 1200

ATTTGCCTTC AACCCAGAAG ACGAAAGCAT TGATAGCCCA CACTTGGCAC CTGCTAACTC 1260

AGTTATGTAC ACAGGAACAC ACGATAACAA TACGGTTCTT GGTTGGTACC GTAATGAGAT 1320

TGATGATGCG ACTCGTGAGT ACATGGCTCG TTACACGAAC CGTAAAGAAT ACGAAACAGT 1380

GGTACACGCT ATGCTTCGTA CAGTATTTTC ATCAGTTAGC TTTATGGCAA TTGCAACTAT 1440

GCAAGATTTA CTAGAATTGG ATGAGGCAGC TCGTATGAAC TTCCCATCTA CCCTTGGTGG 1500

AAACTGGTCT TGGCGTATGA CTGAAGATCA ATTGACACCA GCTGTCGAGG AAGGTTTGCT 1560

TGACTTGACA ACAATTTATC GCCGAATTAA TGAAAATTTG GTAGATTTAA AGAAATAAGA 1620

CAATAATCAG GAGACAACTA AACATGTTAT CACTACAAGA ATTTGTACAA AATCGTTACA 1680

ATAAAACCAT TGCAGAATGT AGCAATGAAG AGCTTTACCT TGCTCTTCTT AACTACAGCA 1740

AGCTTGCAAG CAGCCAAAAA CCAGTCAACA CTGGTAAGAA AAAAGTTTAC TACATCTCAG 1800

CTGAGTTCTT GATTGGTAAA CTCTTGTCAA ACAACTTGAT TAACCTTGGT CTTTACGACG 1860

ATGTTAAAAA AGAACTTGCA GCTGCAGGTA AAGACTTGAT CGAAGTTGAA GAAGTTGAAT 1920

TGGAACCATC TCTTGGTAAT GGTGGTTTGG GACGTTTGGC TGCCTGCTTT ATCGACTCAA 1980

TTGCTACTCT TGGTTTGAAT GGTGACGGTG TTGGTCTTAA CTACCACTTT GGTCTTTTCC 2040

AACAAGTTCT TAAAAACAAC CAACAAGAAA CAATTCCAAA TGCATGGTTG ACAGAGCAAA 2100

ACTGGTTGGT TCGCTCAAGC CGTAGCTACC AAGTACCATT TGCAGACTTT ACTTTGACAT 2160

CAACTCTTTA CGATATTGAT GTTACTGGTT ATGAAACAGC GACTAAAAAC CGCTTGCGTT 2220

TGTTTGACTT GGATTCAGTT GATTCTTCTA TTATTAAAGA TGGTATCAAC TTTGACAAGA 2280

CAGATATCGC TCGCAACTTA ACTCTCTTCC TTTACCCAGA TGATAGTGAC CGTCAAGGTG 2340

AATTGCTCCG TATCTTCCAA CAATACTTCA TGGTTTCAAA CGGTGCGCAA TTGATCATCG 2400

ACGAAGCAAT CGAAAAAGGA AGCAACTTGC ATGACCTTGC TGACTACGCA GTTGTCCAAA 2460

TCAACGATAC TCACCCATCA ATGGTGATTC CTGAATTGAT TCGTCTTTTG ACTGCACGTG 2520

GTATCGATCT TGACGAAGCA ATCTCAATTG TTCGTAGCAT GACTGCCTAC ACTAACCACA 2580

CAATCCTTGC TGAAGCGCTT GAAAAATGGC CTCTTGAATT CTTGCAAGAA GTGGTTCCTC 2640

ACTTGGTACC AATCATCGAA GAATTGGACC GTCGTGTGAA GGCAGAGTAC AAAGATCCAG 2700

CTGTTCAAAT CATCGATGAG AGCGGACGTG TTCACATGGC TCACATGGAT ATCCACTACG 2760

GATACAGTGT TAACGGGGTT GCAGCACTCC ATACTGAAAT CTTGAAAAAT TCTGAGTTGA 2820

AAGCCTTCTA CGACCTTTAC CCAGAAAAGT TCAACAACAA AACAAACGGT ATCACTTTCC 2880

GTCGTTGGCT TATGCATGCT AACCCAAGAT TGTCTCACTA CTTGGATGAG ATTCTTGGAG 2940

ATGGTTGGCA CCATGAAGCA GATGAGCTTG AAAAACTTTT GTCTTATGAA GACAAAGCAG 3000

TTGTCAAAGA AAAATTGGAA AGCATCAAGG CTCACAACAA ACGTAAATTG GCTCGTCACT 3060

TGAAAGAACA CCAAGGTGTG GAAATCAATC CAAATTCTAT CTTTGATATC CAAATCAAAC 3120

GTCTTCACGA GTACAAACGC CAACAAATGA ACGCTTTGTA CGTGATCCAC AAATACCTTG 3180

ACATCAAAGC TGGTAACATC CCTGCTCGTC CAATCACAAT CTTCTTTGGT GGTAAAGCAG 3240

CTCCAGCCTA CACAATCGCT CAAGACATTA TCCATTTAAT CCTTTGCATG TCAGAAGTTA 3300

TTGCTAACGA TCCAGCAGTA GCTCCACACT TGCAAGTAGT TATGGTTGAA AACTACAACG 3360

TTACTGCAGC AAGTTTCCTT ATCCCAGCAT GTGATATCTC AGAACAAATC TCACTTGCTT 3420

CTAAAGAAGC TTCAGGTACT GGTAACATGA AATTCATGTT GAACGGAGCT TTGACACTTG 3480

GTACTATGGA CGGTGCTAAC GTGGAAATCG CTGAGTTGGT TGGAGAAGAA AACATCTACA 3540

TCTTCGGTGA AGATTCAGAA ACTGTTATCG ACCTTTACGC AAAAGCAGCT TACAAATCAA 3600

GCGAATTCTA CGCTCGTGAA GCTATCAAAC CATTGGTTGA CTTCATCGTT AGTGATGCAG 3660

TTCTTGCAGC TGGAAACAAA GAGCGCTTGG AACGTTTTTA CAATGAATTG ATCAACAAAG 3720

ACTGGTTCAT GACTCTTCTT GATTTGGAAG ACTACATCAA AGTCAAAGAG CAAATGCTTG 3780

CTGACTACGA AGACCGTGAC GCATGGTTGG ATAAAGTCAT CGTTAACATT TCTAAAGCAG 3840

GATTCTTCTC ATCTGACCGT ACAATCGCTC AGTATAACGA AGACATCTGG CACTTGAACT 3900

AATACTCTTC GAAAATCTCT TCAAACCACG TCAGCTTTAT CTGCAACCTC AAAGCAGTGC 3960

TTTGAGCAAC TGCGGCTAGC TTCCTAGTTT GCTCTTTGAT TTTCATTGAG TATAAGATAC 4020

AAATTTATAC TAATACATTT TGTAAAAAAG CGAGTTTCGA TTGAAATTCG CTTTTTTAAT 4080

GATGTAGATT TGGGTCAATC TTGTCTAAAA ATAGGGAAAT CCTAGATACA GTGAAGGCTT 4140

TAAATGCTGG TTTTTACTGT CCTCAGCCTT ATATTTTTTC GTAGTTGGTT ACCTCATATC 4200

TATTATATTC GCTTACATAA AGTATTATAA TATAATTGTA GGAAAGAAGG TGTTTTTATG 4260

ATATACACAC TTAAATTGGT GTTGTTTATT ACCTTTCTTG TAATAAGCTT GTTACCTGAT 4320

AAGATTTTTG GAAAAAATAA AAAAATTTGG AAAATAGTTT TTGCAATATT GACGGCAGTG 4380

GCAGCATTGT CATTTATGTA CTAAGTTATT TTAAGAATGT AGGGAAATAA ACCCTACATT 4440

CTTTTTAGTT TTTTCTGTTT TCTAAATTCT ATTTATCCAA GCGATTCAAC ATTTCTTGCT 4500

TCTTCGCTTC AAGTTCTGCA CGCTTTTCTT CGATTTCGGC ATGTTTTTTC TCGAGTTCAG 4560

AACAACTTGC ACCATTGCTA AATTCTTTTC GCCATCAGGA GATAGGGTGA GTCGACATGT 4620

CTATTACTCA CCCAAAGCAG TCCTACAAAG CAGGAATTTT CTGTTACTTT TTTGGAAATA 4680

GTAACGTTTA TACAGCTTTG ACACTTCGTA TCAAAGCGCC AAACACACTC CGAGGGGTTT 4740

ACAGAAAGCA GAAAAGGAAT GATCTGGTAT AAGATCATTC CTTTTCyCTC TTTTTCTTTA 4800

AGTAATTATA TACAATGTAC GACGAAGTCG TCATTGCAAT GCTGATCCAC CACCTAAAGG 4860

GAACTTTAAA CAACATTGAT AAGATAAAGA ATATAAACAA CGAAAATACG TTATACCCAA 4920

TTAATTTTAT TGTATATCTC ATGATTAAAA GTTAATCCTT CCGTTGTTAG GAATGGCATC 4980

ATTTTTATCC CATAATTGTG CTAAATAAGT CCCCGGTGAT AATAAATTCA TAGCGAATTC 5040

TAAAGCAACA TCATTTACAA ACCAACTACC TAGATATCTA GAAATTGCTG AACGAATAGC 5100

ACTTTTTGCT GCATGTTTTC CTTTTACTTT AATTAGATTT GCAAGGCCTG CAGTAGTTCC 5160

TCCTAATGCT AAAGCTATTG CAGTATCTAA TAGAGCACCC ATTTGATTAA CTGTAATACC 5220

TTGCCAAACT GCTCTAAATG GAGAGTATGT AGGTGGGATT GTATAATCGC CTTGTAATTG 5280

TCGGTTAATT ACTTCTTTGA TCCATTGTTG TGAGACGTCT GGATGAAAAG ATTGGATTTC 5340

GTTTGCAAGT GTATTGATTT GTTCTTCTGT TAGAGAAGTG ACAGGTTGAA GTTCCATATT 5400

TGTTTCAATT TGTGATACTT GTTCAGAAGC GTATACAGCT GAAACACTTG GAATCGCTGA 5460

TACAATTAAC ACAATTGACG TCAAAAAAAC CGAAATAAAT TTCATTAATT TGTTCATGAG 5520

CTTTTCTCCT TTTTATTTGC ATCTGCTTAC ATTTTATCAT ATACTGTTAT TATAGTCAAA 5580

AAAATATGCT ATTATGTTAA AAAAATATTT TTCAAAATAT AAATGGACGG ATTTATTTTG 5640

GATTTTATTT GTTATTTTGA CCTGCCTCTA TATTGGTAAC CATGATTTGT TTACTCTCAA 5700

TCATCAAGAA TTCTCTTTTC GTGGTAGCGT TTGGGGTCTG GTACTGGCCT TATATCACTT 5760

ACTATTCATT GATAAGTTTG TTATATCGAA TCGAAAATAA AGATTAGAGC TATGCTTGAC 5820

TGTGTACTTT TAGGATTTAT TTTGGAGGAA GATTTTGTCT CTATTATTTA TTATTTTAAA 5880

TTTATTTATT TTGTATAAGA TCTATTCTTT 5910

(2) INFORMATION FOR SEQ ID NO: 166:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5406 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

GGCATAGCGA CTCATTTTTT CAACTGTCCA GGCTGGATAC CAGACTAATT TAACCTCAGT 60

ATCCGTTACT TCTGGAACCT CTATCATAGC ATCATAAATC TGGTCTGTCA AAAGGTCTGC 120

TAAGGGACAA CCCATAGTTG TCAAAGTCAT GTCAATCTCT GTTTGCCCTG TGTCACCGTC 180

AAAACGAATC TCATAGATCA AACCAAGATT GACAATATCG ATTCCCAACT CAGGGTCGAT 240

GACTTCTTCC AAGGCTGTTA AAATCCGTGT TTTGATGTTT TCAATTTGCT CTTCTGTATA 300

AGCCATATTT TCCTCACTCT TAGTCTTCAA TAAAATCACG AAGCGGTTTG CTACGACTTG 360

GTTGGCGTAG TTTTCTCAAA GCCTTTGCTT CAATCTGACG GATACGCTCA CGAGTTACGT 420

TAAAGACTTT CCCCACATCT TCAAGTGTGC GCATTTTTCC ATCATCTAGT CCAAAACGTA 480

GACGCAGAAC ATTTTCTTCA CGGTCTGTAA GAGTATCTAA GATTTCATCC AATTGCTCAC 540

GCAAGACGAT ACGAGTCGTA TAATCCACTG GATTTTCAAT CACTTCATCT TCGATAAAGT 600

CTCCAAGGTG GCTATCGTCC TCTTCACCGA TAGGAGTTTC AAGAGATACT GGTTCTTGGG 660

CAATCTTCAA GATTTCACGA ACCTTATCAG GTGTCATATC CATTCGTTCA GCAATCTGTT 720

CTGGTGTCGG ATCTTGCCCC AATTCTTGAA GGAGATTCCG CTGTTCACGA ACCAATTTAT 780

TGATAGTTTC AACCATGTGA ACTGGGATAC GGATGGTACG AGCTTGGTCC GCAATAGCAC 840

GAGTGATAGC CTGACGAATC CACCAAGTTG CATAAGTTGA AAACTTGAAC CCTTTAGAAT 900

AGTCAAACTT GTCAACCGCC TTCATCAAGC CCATATTTCC TTCTTGAATC AAGTCAAGGA 960

ACTGCATACC ACGACCGACA TAGCGTTTGG CAATGGAAAC AACCAAACGA AGATTGGCTT 1020

CCGCAAGACG TTGTTTGGCT TCGATATCAC CAGCTTCAAC AGCCAGTGCC AACTCTTTCT 1080

CCTCTTCATT GGTCAAGAGA GGAACGACCC CTATTTCTTT CAAGTACATA CGGACAGGGT 1140

CATTGACCTT AGCAGAAGTT GACCCAATCA AGTCCTCATC GCTGAGTTCT GGTTCTTCTT 1200

CATTGCTGAG AACACGCGCA CTTGGATTTC CTTCGTTATC TGTGATAGAA ATGCCTGCAT 1260

CCTGAATCCG TTGCAAGAGA TCTTCAATCC CATCAGCGTC CAAGGTAAAA GGAATAACCA 1320

GACTTGCATT GATTTCATCA TCTGTTGCTG TCCCTTTTTG CTTATGATTA CGGATAAATT 1380

CTGCTACCTG TACGTCAAAT GTTGTTACTT CTTTTTGTTT TGTTGCCATT ATTACTCCAT 1440

TCTTCTCTTT TGGGAAATTA AACGTTCCAA TTCTTCTAGG GCTGTATCTG TATCTCCTAC 1500

ATGGCTAGCT TCCTGCACCT TCTTTTTGAT TCTCATATTG TCCTGATTCA AGAGAGCCTT 1560

GTTTCGAGTC ATCTCTACTT CACTAAGTTC CTGCGGCGAT ATCTCAGCAG GCAAATCCTG 1620

AGCTAAAACT TGGTACCAAG CTCTTTCAAC TTCCTCTGTC TGCTCTGCTA AAACTTCTGG 1680

AGGAAGATTT CCATACTGGC CAAGCAAGTC ATATAAGACC TGAAATTCAG GTGTAGCAAA 1740

TGCAAAGTCT TCTCGCAAAC GGTAATCGTT CAAAACAAGA GGGGATTCCA TCATCCGATA 1800

GAGTAGATGG GCTTCTGCCC TCATAATAGC CGATAACTGC TTGGTGACAG GCATGGTGAT 1860

TGGCGTCGGT CTGGAAATTC CTTCCATGCG ATTCTGCCTT TGCACCTGAC GACTCTCATT 1920

AACAATCTGC TCAATCTGGG TATAATCAAA GGACGCCAGA CTGTCAGCTA AAATATGAAT 1980

ATAGCTGTTT TGAGCAGCGA TGGACTTTTC TTGAACAATC AAGGGAGCTA TTTTTTCAAG 2040

AAACTCAATC TGAGCCTGCA GATTTTCACT GTTTTCAGGT TTGTACTGAT GAATGTAGAA 2100

CTCAATCGGA CTAATACGAG TTTTCGTTAA TAGATAGGCC AAGTCTTCTG GACCATTTTT 2160

TTGTAGATAC TCATCAGGAT CCAAGTTATC AGGCATGCTG ACGATTTGCA CAGGCATATC 2220

ACCAATTTCA TCCAATGCTT TCAATGTCGC GGCTTGCCCA GCCTTATCTC CATCGTAAAC 2280

AAGAACCAAT TTCTTGGTTA ACCTTTTCAG ATGCTCAACA TGCTCTCGAC TCAAGGCTGT 2340

TCCCATCGAC GCCACAGCAT TTTCGATTCC AGCCCGATAG GCTGCAATAA CATCCATGAA 2400

TCCTTCCATC AGGTAAATCT CACTAGCTTT TCCAGAAGAT CTTTTTGCCC TATCCATATG 2460

ATATAATTCG TAACTTTTGT TAAAAATTGC AGTCGATCGG CTGTTTTTAT ACTTAGAAGT 2520

TTGTGAATCC GTTTTTTGCC AGATACGACC TGAGAAGGCA ATGACCTTTC CTTGGTCATT 2580

TGTCAGGGGA AACATAATGC GATTGTGAAA GGTGTCTACA AATTGATTGG CATCCGAGAG 2640

ATAAAACAGG CCTGAATCCA GTAAATCCTC TTCACGATAC TGATCAGACA AACGTTGATA 2700

GAGATAGTTT CGTTCTGGAG GTGCTAAACC AATCCAAAAA TGTTTAAGCA CTTCATCTGT 2760

CAACCCCCGC TGATAAAGGT AATTTCTGGC CTCTTCGCCC ATAGTCGTTG TCATGAGAAT 2820

AGCATGGTAA AATTTGGCTG CATCTTCGTG CATATCATAA AGAGCTTGGT GAGGTGAGGC 2880

TGACTTCTGC TCACTATAAA GCGGTTTTTC AACCTCAATT CCAACACGCT GACCTAAGAT 2940

TTGGACTGCT TCTATAAAGG GAACCCCTTG GTACTCCTCG ATGAACTTAA AGACATCACC 3000

TGAGCGACCA CAACCGAAAC AGTGATAAAA CTGCTTGTCC TCTACAACAT TGAAAGATGG 3060

TGTTTTTTCA CCATGAAAAG GACAGAGCCC TAGATAGTTC CGTCCTGCCT TTTGTAAAGA 3120

AATCACATCT CCTATGACTT CCACAATGTT GGCATTGTTT TTGATTTCTT CAATGACTTG 3180

TTTGTCAACC ATACACAATA CCTCCATGTT ATCATAGTTT ACTTTATATA GTATACTTTA 3240

TTTCAGAAAA AAAGTAAACC ATTTCACTCA TTTTCCCTAC TTTATTCAAA GAGTTGATAA 3300

TAATCAGAGA TTTTCATTTT TGCTTTTTCT TCTTGGTTTA AATCTTGGAT AATTCGTCCT 3360

TCTTTCATGA CAATCAAGCG ATTGCCGTAT TTGAGAGCAT CTTCCATATG ATGAGTAATC 3420

ATAAGGGCTG TTAGCTGATC TTTCTTAACA AATTCATCTG TCAATTCCAT CAAAGCAACA 3480

CTAGTCTTTG GATCCAGGGC AGCAGTATGC TCATCTAACA GGAGTAATTC AGGTCGCTTC 3540

AAGGTTGCCA TCAAGAGACT CAAAGCCTGT CTTTGTCCAC CTGATAAGAA CTCAATCGGT 3600

GTATTCAAGT GTTTCTCAAG ACCATTTCCT ACTTTTTCAA TGGTTGCCTG AAATTCATCC 3660

TTATAGCTAG TCAAGCGTCG TGGTAACAAT CCACGCTTTT CACCACGAAA CTTGGCGATT 3720

AAAAGATTTT CAGCGACCGT CATACGGGGA GCTGTCCCCA TCTTTGGATC TTGGAAGACA 3780

CGAGACAGGT ACTTGGCACG CTTCTCGGGT GAAAACTTAG TGAGATCTTC ACCTAAAATA 3840

CGGATAGTTC CACTAGTTAG TGATAAGGTC CCTGCTATAG TGTTAAAGAG AGTTGATTTT 3900

CCAGCACCAT TTCCGCCCAA AATCGTGATA AAGTCCCGTT CAAAAATTTC TAAGGAAACA 3960

TCATTTAAAA TAATCTTTTC TTCATCAAAG CCATTTTTAA CGATTTTGGT TGCATTTTTT 4020

AATTCTACAA TTGCTGTCAT TTGCTTAACT TGGCTCCTTT CAAGATTGTT TGCTTAAATG 4080

TTGGAATCAT GAGGCAGACT GCTAAAATCA AGGCACTGTA TAAACGAAGG TAACTTGTAT 4140

TAAAGCCAAG TGCGATAACT GCCCACACTA AAAATTGATA AGCGATAGAA CCTACAACGA 4200

TAGTAACCAA ACGCTCTGCC AAGCTCAAAC TCTTGAAAAT AACTTCTCCA ATAATCAAAC 4260

TTGCAAGCCC CACAACGATA ACCCCGATCC CTCGAGACAC ATCGGCATAA CCTTCTTGCT 4320

GAGCAATGAG GGCACCTGCA AGGGCAATCA CACCATTTGA TAAGACCAAG CCCATGAGCT 4380

CCATGCGTCC AGTATGAATC CCGAAACTTC TAGCCATATC AGGATTATCC CCTGTAGCAA 4440

TATAGGCTTG TCCGAGTTTA GTGTCCAAGA AAAAGAGCAT GAGAGCAATA ACAATACTCA 4500

CAAAGATGAG ACCTGTCAAG AGTTGATTCA AATCCGAATC AAAAGGCAAA ACATCCTGAA 4560

TTTGCTTGGT TCCAAGCAGG CCTAAATTCG CACGTCCCAT AATCAAGAGC ATGATTGAGT 4620

GACAAGAAGT CATCACCAAA ATCCCTGAGA GCAAGGTTGG GATCTTCCCT TTTGTATAAA 4680

GAAGGCCTGC TGCCATTCCA GCCAAACAAC CTGCTCCTAC AGCAACAAGT GTCGCTAAAA 4740

ATGGGTTCAC GCCTTTGGTT ATCAAAGTGA CAGCAACAGC TCCCCCAAGA GGGAAGGAAC 4800

CTTCTGTCGT CATATCTGGA AAGTTTAAAA TCCTAAATGT CATAAAGATT CCCAGACCTA 4860

GAATAGCCCA GACAAATCCT TGAGAAATAA TGGAAACAAT CATATTTTAT TTAATCCTTT 4920

CTATATTCAT CTTTTTAAAA AATGGGAAGA GTCTCCTCCT CCCTACCTTA TTTATTCGAT 4980

GACTTGTCCT GCTTCTTTGA GAACAGACTC AGGAATAGTA ATACCTAGTT CTTGTGCTAT 5040

TTTTTTATTG ATGACTGACT TACCAGTTGA AAAGACATTG ACTGGGGTAT CGGCTGGTTT 5100

TGCACCTTTC AAGACTTGCA CAATCATTTT ACCTGTTGCC ACACCAAGGT CATGTTGGTC 5160

AATTACAACT GATGCCAAAC CACCTACTTC TACCATAGCT GTCGCACTGG GATAAATTGG 5220

TTTCTTAGAA CTTTGATTGC TAGAGACAAC CGTTGGAAAT CCTGATGCAA TGGTGTTATC 5280

AATTGGAACC CAAATAGCAT CTACCTTGCT AGTCATAACA GTGACAGTTG AGGCAATTTC 5340

ATTTGTTGAA GGAACTGCAA ATGTTTCCAC TGTCAGACCT GCCTTTTCAG CATAAGCCTT 5400

AAATTC 5406

(2) INFORMATION FOR SEQ ID NO: 167:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9711 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

CAGCTTGCTC TTACTATTAT AGCAGATGTT ATAGCTGGAA TTATCTTGTA TTTCGTCTGC 60

AAATGGCTAG ATGGTAAGAA GTAGACCGAA TGACTAGCCT ATAAACACCC GTTAAATCGC 120

TAAGATACGT CAAAAAAGCC CTTAACTATG GCACTAGTTA GGGGCTTTGG TGTTCTAATG 180

AACCTTATAC ACTAACTACA TTCTAGCATA TAAGCCCAGA TATTTCAAGA GTTTTATTTA 240

TTGTTTAAAG TTCTGAAAGG TCTATAATGA AGTTAGCCAT CTAGTATCAA AAAACCGACT 300

AGCTCTTATG AACTAGTCGA TTTCTCATCA ATGCGCCAAC ATTTCTTGGG CGATTTCTTG 360

GCCAGATAGG TTATCTGGGT AGTAGGTTGG CCAGTTGTCC ATTTCTTCAA AGAGGGCTTC 420

TTGGCTTGTG CCTCCAAAGA AGATATGGAA ATGTTCTGCC TTAACTGGGG CAACATTGTG 480

GTCACTAAAC TGAACATACT TGAATTGTCC AGCGTCAGCA TCTGTGGCTT CAAAGAGGAA 540

ACGCACGCCA CGATTGCCTT TCTTGTAAGT CAAAATTTTC TTACCGACAT ACTTGTAAGT 600

GTATTTCTTG CTTTGTCCAC CTTGAACAAA TTCCATAGTA TTATCAGTAA TGTTAATCTT 660

AGTCACATCT GTATGATAGC CTTTTGTATA GTAAGCCTTG TACTCAGCCT GGGTCATCTT 720

ACCAGTCAAC TTAGCCTTGT AGTCAAAGAC TTGGTCAAAC GTGCCGTCTT CAAGGAAAGG 780

ATAAACTGAT TGCCAGTTAC CTGCATAGTC ACTCAAGGTG CGGTCCTTGA CAGCTGCATC 840

CTCGAAGTAA CCATTTTGGA CTGTCTTGGT ATCCTCTGCC TTTTCAGGTT CAATTGCTGG 900

GCCTTCTTGG TCTGTTGTTT GTTTCAAAGC CTTGAGGTTT TTCTCCATCA CGGAAATGTA 960

GTTTTCTCCA GCCTTGGTGT CCTCTTCTGT CAGACTTTCT AAAGGATTGA GGACATCAGT 1020

TTTGACACCT GCTTCTTTTG AAAGTGTGTT AGCAAGGGCT TGTGAGGCAT TTCTTCAAAA 1080

TAGATATAGG CGATTTTATT TTTCTTGACA TACTCTGTCA ATTCTGCCAA GCGAGCAGCT 1140

GATGGCTCTG CATCTGGAGA AAGTCCTGAG ATTGCGACTT GTTTGAGTCC ATAGTCCAAG 1200

GCAAGATAGT TAAAGGCTGC GTGTTGAGTC ACAAAGCTCT TTTGTTTTGC TTGAGACAAA 1260

CCTTCTGCGT AAGCCTTATC CAAGGCTTGC AATTTTTCGA TATAGGCAGC TGCATTCTTC 1320

TCAAAGGTCT CTTTTTTATC AGGATAATCT GCTGACAAGC TGTCGCGGAT GTGCTCTACT 1380

AGTTTAATGG CACGAACTGG TGATAACCAA ACATGGGGGT CAAACTCATG GTGATGACCT 1440

TCTTCTCCAT GGTCATGGTC TCCCTCTTCT TCCTCGCCAC CTGGCAAGAG CAACATATCG 1500

CCTGTCGCCT TGATGGTTTT CACTTTTTTC TTATCCAAGG TATCTAGCAA TTTAGGTACC 1560

CATGTTTCCA TGTTTTCATT TTCATAAACG AAGGTATCTG CATCTTGGAT TTTGGCAACT 1620

GCCTTGGCAG ATGGTTCGTA TTCATGAGGT TCTGTCCCAG CACCGATTAG GAGTTCTACA 1680

TTAGCCGTAT CTCCTGCGAC TTGCTTGGTA AATTCATAGA CAGGGTAAAA GGTTGTCACG 1740

ATATTGAGTT TACCATCTGC CTGTTTTTGA TTGGAACAAG CCACTAAAAA CAAGGCACAT 1800

AGACTGGCTA GTAATAAGCT AATTTTTTTC ACGTTCGTCT CCTATTTGAT AAAACGTCTT 1860

ACTAAACTGA TTAGTATAAA GACAGTTACA AAAATAATGG TAATACTTGC ACTTGCAGGT 1920

GTTTCTGCAT AGTAGGAAAT GTAAAGTCCT GCTACCATTC CCAAAAAGCC AATCGCACTG 1980

GCAAGCAGCA TAACCGATTT AAAGTTTTTC CCCAGACGCA GGGCAATACT AGCTGGCAAG 2040

ACCATAATGG TCGATACCAG AAGAGCTCCT GCTGCAGGAA TCATAAGGGC AATAGCCACC 2100

CCTGTCACCA TGTTAAAAAG AATGGACATG GTACGAACTG GCAAGCCATC CACAAAGGCC 2160

GTATCTTCGT CAAAAGTTAA GATATACATA GGACGAAGAA AGAGAAAGGT CAAAATCAAA 2220

ACAACCGCCG CAATGACAAA GAGGGAAATG ACCTGTTCTT CACTGATAGT CACGATCGAA 2280

CCAAAGAGAT ATTGGTCCAA ACTCATTGAA CTCGAGCTTT TACCCTTGCT CATGACAATC 2340

AGAGAAACAG CCAGACCTGT TGACATGAGG ATAGCTGTCC CGATTTCCAT AAAGCTCTTG 2400

TAAACCGTAC GGAGATACTC CAGAAAGACC GCCGCAATCA AGACAATGGC AATAGTAGAA 2460

ACAGTTGGAG AAATCCCCAA AACCAGACCA AAGGCTACAC CTGAAAGTGA GACGTGGCTA 2520

AGGGTATCAC TCATCAAACT CTGACGACGC AAGATGAGGA AGGTTCCCAA TACCGGTGAG 2580

AAAAGACTCA TAGCAATAAC CGCCAAAAAG GCGCGTTGTA TAAAGTCGTA AGATAATAAA 2640

CTAAGCATGG CCCACCTCCT GGCCATTCTC ATGAACATTG AAACAACGCC ATGGCGAGTC 2700

TTGGTTACGG ACTAGATGAA TATTGCGATC CGCATAATCC TTAACTTCTT CAGGGTCATG 2760

GGTAATCATC AAAACAGCCT TGCCATGATG ATGGGCGCTG TGGTGCATGA GTTCGTAAAA 2820

TTCATTTTTA CTTCCTGCAT CCATCCCCGT TGTCGGCTCG TCTAGGATAA ACACATCAGG 2880

GTCAGAAGCA AACATACGCG CAATTACCGC TCGCTGCTTT TGTCCCCCAG ATAGAGACCC 2940

CAAGCGTTTG TCTCGATGTT CCCACATGCC AACTGAGTCC AGACTAGCCT TGATATGCTC 3000

CTCATCATGA GCATTCAAAC GACGGAACCA GCCTTTTCTC GGATAGCGAC CCGACTTGAC 3060

AAATTCATAG ACCGTACTTG GAAAACCAGC ATTAAAACTG GCAATTTGTT GAGGAAGATA 3120

GGCTATTCTC AATTTCTTAC CTTGCGTATT TGTCTTTGAA ATAGCCACCT TTCCAATGCG 3180

TGGTTGCAGA ATTCCAAGAC TAGCCTTGAT GAGCGTCGTC TTAGCCGCTC CATTTTCCCC 3240

AGTCAAGGTA ACAAATTCCC CACTATCAAC ACAATAATTG ATATGTTCAA GAACAGGCTC 3300

CTTATCATAA TAGAAGGACA AATCCTCTAC CGTAATATAT CTCATTATTT GATTTCTCCT 3360

ACTAAAGCAG TCAAAAACCG CTGAATCACT TTTTGTTCAT TTGGAGTAAA CTGAGTCGCC 3420

ACTTGTTCAT AGGTTAAAAG TGTATGCTCA TGGTGATGGT GGTGCTCCTC AGCGATTGGA 3480

CGAGCCAAGT CAGTCAACTG ATAAAAAATC ACACGCGCAT CTTTAGAATC TTTAGATGTT 3540

TCCAACATCC CTTCCTTGAC CAAAGACTTA ATGGCCTTGG TAACTGCCGC CTGACTGACA 3600

TTGAGACGAC GGGCCAATTC TGAATTTGTT AAAGATTCCT CTGACAAGAG CATAAGGATA 3660

TGCTCCTGAG TATTGGTCAG GGCCACCTCG CTAGTGCAAT GACCTATTAG GATTTCATGC 3720

TGATTTTCCG CCTGCAAAAT CACCTCATTC AAAAAAGCAT TGATATCCTT TGCTAGCTGT 3780

CTCATATCTG ACTCCTTTCC TTTTAGACTT CTCTTTTTTA AGAGAAAAAT ACTATTCTTT 3840

GACATTTTGT TTACCAGTTA ATTATATCAC AAGCAAAAAA AGAGTCAAGA AAAAACGTGA 3900

AAACTAGTTT CATTCTTGAA CTCTTCTATA TTATATTATC TATTGAAATT CTTTGACATC 3960

TCCATCATAA GTCGCCCAAT CTTTGCTGAA AAAGCGCTCA TTCAGATGGT AAGTCGGAGC 4020

TGGTGTGGGA TTGGATAGGA AAGGATCAAC TGCCTTGTCA AAAGCCAACC AACCCAACCA 4080

ACCAAGGTGA ATGGTGTCCT TCATAAAGAA AGGCTCCCCG CCGTCCTTAG AAAAATCTGC 4140

TATATTGGTA AAACCTTGAC TTTCTAACTG GTAGCGAATC TTCTGCACCG TTTGTTGGTA 4200

CATATCCTCT CGTAGACCAG CATAGTTCAT CCATTTTTTA TTAACAGGTG GAATGATAAA 4260

AATCGGGTTT ACCTTAGATT TAGAAAACTG TGTTAAAACC AACTGCAAGT CATTATACTC 4320

TGGCGACTTG AGATAGGTAA AGCTTTTCTG AGAATCCTTT AATTTCTTCA AATCCTTCTT 4380

GATCTGCTCA TTATAGAAAT AATTTTCCAT TCCCATCTCA TTATTGGAAG TATTTTTTTC 4440

AGCATCTGCT TTGACAACAT CTTCTATTGC CTGATAAGAA AACTGGTCTG GCAAGATTTT 4500

TAAATACTTA GCTACATGCT TATCGTAGTT AACATAGCCT CTAACCGAAA ACTGACCAAA 4560

AAAGGAAGCT TGGCGTTCAT TAAAACGAGC CAATAATTCA ATCATTTCAT TGTCTGCTGT 4620

CGACAATTCT TCTTTACTTG CCAACTTCTG AACCAGGTCC TTCATAGCTA CGTTTGGGAA 4680

CTGTTGCAGT AAGCGAGTCG CTGCATATTG ACTAGCCTGA TCCCCAGATT GATGTTTCAG 4740

AAAACTAGTC AACTGGTCTC CATTAAAATA CTGCTGGAAG GCTGCTGGAT CATAGCCATT 4800

TTTACTGAAC CACTGAGGTG AGATAACATA CACAACTTGT TTATTCTCCA GCTGTGGTAA 4860

CATCTGTTGC ATTCCAAAAT ATTGGTTAAG CGATGCAGCT CCCCCCTGTC CTAAAAGATA 4920

AGGACGGTAG GAACGATTGT ATTTCTCAGC TAATACCGCA GGATGAGCAC CGTCAAAACG 4980

AAGCCATTCA CTAGAGCCAA AGAAGGGAAC AAAACGCACA TTTGGATCAG ATAGTGCTCT 5040

GACTTTTTGA CTTCGCTCCT TAAAACTATC GATAGTAGTA GCCACTGCTG AACGCTTTTC 5100

AGCTCCTAGA TTATGATGCA TCTCAGTAGG ATAAAAGAAA ATGAGCAGAA AAACCAACAA 5160

ACCAGCGATC AAGACCGGTC CGAAGATCAT CCATAAGCGT TTAAGCATTT TGTAGCTCCA 5220

CAATACCAGC TATGATTTTA TTAGCTGTAT TCCAGTCGTC ACGACCAAAC TCTGTTACAG 5280

GGACACGAAT GTCAAAACGG TTCTCAATCT CCACAATCAA CTCAACCGTT CCCATACTAT 5340

CCAAGACACC TGCATCAAAA AGATCTTCAT CCATCATGTC AGAAACATCT TCCATAAACA 5400

ACTCATCAAT AATTTCAATA ACTTCTGATT TGATATCCAT ATTTTATTTC CTTTTATTTT 5460

TTAAACCATA GATTATTCAA GAATCCAGAA AAGATTAAGA ATGACAACAT GACAACATGG 5520

AAAGTGACAA CCATGCCAAG CAACTGAATC CAGCGATTCT CAGGTAGGGC AGCCTTCCCT 5580

GCTTTTTTCC GTTCCTTATT GAGCGTTTTT TTCTTGCGAA CCCAGGCATC ATTGATGACC 5640

AAGCCTAGTC CATGAAAGAG TCCATAGGCG ATATAGTACC AGGTCACACC ATGCCAAAAT 5700

CCCATAATCA GCATATTTAC AATGTAGGCC ATGCTTGAGG TTACATTACG ATTTTTAAAG 5760

ACTTTCTTTC TGGTTAACAC CATCACCATT CGCATAAAGA CAAAGTCACG GAACCAGAAG 5820

GACAGACTCA TATGCCAGCG ATTCCAAAAC TCCTTTAAAT CCCTTGATAA AAAGGGCTTG 5880

TTAAAGTTGA TAGGGCTACG GATTCCCATC AAGTTTGAGA TGGCCAAAGC AAACATAGAA 5940

TAACCTGCAA AGTCAAAGAA GAGTTCCAGA CCAAAAGTAT ACATAACTGC CAAGGCATAG 6000

AGATTAAAGA AGCCACCTGA CTGCAAGGCT AAATTCTTCA GAGGAGGTAG TAAGGTCTCT 6060

CCTAAAACAT GAGCTAGGAT AAACTTATAC AAAAAGCCCC ACATGATATA GCGGACAGAT 6120

TCATCCAGCA TATCCATCAA CTCATCTCGC TCAGGAATAG CCTGATAATT TTCATTAAAT 6180

CGCTTAAAGC GATCGATTGG ACCACTCGAG AAAGTTGGCA TGAAGAGAAG GAAACGGAGG 6240

AATTCCCAGA GGGTAAAATC CTTAATCACT CCATCTCTCA GCTCGATGAC AATTCCAACC 6300

GAACGAAAGG TCAGGTAAGA AATTCCCAAG AACCCAAGCA AAGACTGCGT TCCATTGATA 6360

GCTGGTTGCA CCTTGACAAA GATAATCGGA AGTAGGGACA GAAAACTAAC TAAGTAGAAG 6420

ACCCACTTGC CATCCTTGCT TTTTCGATAA TGCTTGTAGA AAAGCAGGAG CAATATTTCC 6480

CAGCAAAGGT AAATACCCAA GGCAGCTAGT TGATTGGTCT TTCCACCCAC CAACATGGTG 6540

ACAATAAAGA AGAGACTTAC CAACACTTCA TACCAGGCAA AGCGTTTCTT GAAAAAGAGA 6600

CCTATAAAGA TGGGCAAGGT TGCAGCAATC ACATAAACAA AATACTGAGG ATTGCCATAT 6660

GGCTCTAAAT GAGGAAGCTG TTGAAAAAAC TCCATCATCT CTTATTCACC TCGTTAATCA 6720

ATCCTTTGAT GTCAATCTTT CCATTTGGAG TTAGTGGCAA ACTGTCTCGG TAAAGGAATT 6780

TAGATGGCAT CATATAGGAC ATCATGATGT CTGTCAGGTC TTCCTTGATG GCCTTGGTAA 6840

TATCGATATC TCGCTCAAAC TGCTCACGAA CACCGTCTTT TAAGATGACA TAAGCCAATA 6900

GATTTTGTAC CTTGTGGTCC TTGTTATAGC GCGGTACTGC GACAGCAGAT TCGATAAAGC 6960

GAGACTTGTT GAGGTTTTGA GAGACATCTT CTAACTCAAT GCGGTAACCG TTAAACTTAA 7020

TCTGGAAGTC CATGCGTCCG CCGTAGAGAA GCAAGCCCTC ATCTGTCATG GTTCCCACAT 7080

CGCCTGTGTG ATAGGCTGGC AGATCTTCAA ACTCAAAGAA GGCTTCTGCT GTTTTTTCAG 7140

GATTGTTCAT ATAACCTTTT GAAACAGCTG GCCCAGAAAC AATGATTTCT CCCTGCTCAC 7200

CATTTGGCAG TTTATTTCCT TCCTCGTCAA TGATAAAGGT TGGAGAATCA GCCTTGGTAT 7260

AGCCGATTGG TAGGCGTTTG AGAGTCGCTA ACATCTCGTC TGTCACGGCA ACTGCTGACA 7320

GAGCTACTGT CGCTTCTGTT GGGCCGTAAG CATTGATGAT ACGGGCATTT GGGAAACGCT 7380

CGCGCAGTTT TTGAGCTGTT TTGACCGTCA ATTCTTCACC ATCAAAGTAG AAATGCGTGA 7440

TTCCAGGCAT TTTCTCACTG TTGAAGTATT CAGACAACAT GGCCATATCT GCAAAGGATG 7500

GTGTTGATGT CCAGATAGCG ATTGGCAATG AAAAGATAGC CGCAAAGAGT TGCTTAAAGT 7560

CCTGAGTGAT GACTGAAGGA AGAGTGAAAA GCGTACCACC AAGTGCCAAG GTCGGTGCCC 7620

AATACATGAC AGACAAGTCA AAAGAATAAG GTGGCTGTGC CAGCATTTGC GGACGACTCG 7680

GTGTCGCAAA TTCCTTATCC GTAATCATCC AGTTTGTAAA GCTGAGGAGA TTATCATGTG 7740

AAATCTGCAC TCCCTTAGGC TTACCAGTCG TACCAGAAGT AAAGATAATG TAGTAATTAT 7800

CATCTCCCTT GACTGGATGC GTGATTTCAT AGTTATTCCC TTGGGCAAAG GCTTCTTGAA 7860

CCTGAGCTAG ATTTATCATT GGTGTAGAAA CCTGCTCCAA GGGAAAGGCT GAAATGGCAA 7920

TAATCAAGCT TGGCTCTGCT ACTTCTAAAA TAGCTGAAAC TCGCTCCAAG GCCGAATGGC 7980

TATCAATTGG AATGTAGGCA TGACCTGACT TAGTCAGCGC TACAAAGGTT GCCAACATTT 8040

CATATTCTTG GCCACCAAAA ACAACCACAG GAGACTTCTC AGGCAAGCCT AGTTGGTCAA 8100

TGACTGCAGC CAAACTATCC GAATCAGCCT TTAAATCGCC ATAAGTGTGT TCCTGCCCCA 8160

AAACATTATA GACAGGATAG CTAGGCTGTG TCTGAGCAAA ATGCTCAATG GTTTCAATCA 8220

TATCTGCTAT TGGTTTATTT GACACAATAG GGATTCTCCT TCAAGTTAAA ATTCATTATA 8280

GATAAAGCTT CCTTGACCCT GACCAAGATA GCTAAAGAAG TAAAGCAGCC CTAGAAAGAT 8340

AAGAAAATAC AAGGCTGTCC GACCAAGAAA GAGGTACAAT TCTTTTCTCT GTTTCATCAA 8400

GAAAAACCAT TCATTTCTGT AATTTTTCGC TAAAATAAGA GTGATTCTTA CTAGCTTATT 8460

TTTCTACCAT TGTACCACTT TATATAGTAT CTTTTCAATT GTTTACCGTA TGTTTCCAAT 8520

AGATTTCAGC TTATTTTAAG GATTATACAG TTTTTCTATG TATATTTTCA AATAGAGTGA 8580

TCCTGCTTCA AAACTCCATT TCAGGAGACA ATGAAGTAAA TCTTCCCATA ATAAAACACA 8640

CAATATCAAG TTTTTTCAAC ACCTGATACT ATGCGCTTTT CTGATTTTTA AAGACTTTTT 8700

AACCACTCTC TCATTTAAAA TAATCTCGTC TGATATAAAT TAAAATAGCT TCTATCATCA 8760

GACAAATGGC TGATAGCCAA AAACTGATGC TAATACCAAA ACTCTCAGTA ATATAGCTCA 8820

TTAGCAAAAC AAATACTGAA AATGCTAATG TAGAAATCAC TTCAAGAACG GAATAGACAT 8880

TAACTAAATG ATTTTCCTCT ACTGTTTCCT GAAGAAATAC ACTTTCAGGA ACTTCTTTTA 8940

GTTGCGATAA CATACCAACT AAAGCTGAAA ATAATAAAAA CATCTGTGCG TTTGGAAAAT 9000

ATAGAATAGT CAGTGTCACT ATTTCCATAG CTACAAGAGG AAAAAGAATA CTTTCCCCCC 9060

AAATCATTCA TACCTCTCTC AACTAGATGT AACTTACAAA ACCCCTGACC TCATGAGCCA 9120

CTTTCTTCCT CCTCATGAGG TCAGTTTTAC TTTCTGCTGT TCCAGTATCG TTTTTCCTCG 9180

CTAGATTTCC TCAAAAGGGC AGACTCCTCC CTTGGTGCGT CACACGATTT TTTCATCTCG 9240

ACTGTTCTTT AATGCATCAT TAACGACGCT TTTCTTCTAG GTGGTTCATA AGGAACAGGA 9300

AGATTCAGGT TGACTTTTCT AATCCTAGAA TAAAGTGCTG AAAACAATTC GGAATAGGCA 9360

TAGAGACTAG ACAATTTGAG GAGCTGCTTG CGTCCTGTTC GAACACATTT TCCCACCACG 9420

TGAAGAAAAA GATGGCGGAA GCGTTTGATT GTTAAAGTTT GGAAGTCACC TCCAGCTAGA 9480

TGTTTGAGAA AAAGATAGAG ATTGTAGGCG ATACAGCTCA TCATCATACG AACTTCGTTT 9540

TTGATTAAGG TTGAACTATC CGTTTTATCG CCAAAAAATC CCTCCTTCAT CTCCTTGATG 9600

AAATTCTCGG CTTGACCACG TCCACGATAA AGCTGAAACT GGTCTTGGcT gTTCCACTCG 9660

TCATATTTGT AACGAGAGAA ATAACATCGT AGAACAAGTA TCCTTCTTTT C 9711

(2) INFORMATION FOR SEQ ID NO: 168:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3025 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

CCCCTTTGTC AAAACTGTAA AATTAACGAC TCAACAATTC ATCTTTACAC CAATCTCAAT 60

GGAAAACAAA AACAAATTGA CCTCTGTCAA AACTGCTATA AGATTATCAA AACAGATCCT 120

AACAATAGCC TCTTCAAAGG TATGACGGAT CTGAACAATC GTGACTTCGA TCCCTTTGGT 180

GATTTCTTCA ATGATCTAAA CAATTTCAGA CCTTCTAGCA ATACTCCTCC TATTCCCCCA 240

ACCCAATCAG GTGGAGGTTA CGGTGGAAAC GGCGGTTATG GTTCCCAAAA TCGTGGATCT 300

GCTCAAACTC CGCCACCTAG CCAAGAAAAA GGCCTGCTGG AAGAATTTGG TATTAATGTA 360

ACTGAAATTG CCCGTCGTGG AGACATTGAC CCCGTTATTG GGCGCGACGA TGAGATTATC 420

CGTGTCATCG AGATTCTCAA TCGTAGAACC AAGAATAATC CTGTCCTTAT CGGTGAACCT 480

GGTGTCGGAA AAACGGCCGT TGTCGAAGGT CTAGCTCAGA AAATTGTCGA TGGCGATGTG 540

CCACATAAAC TCCAAGGTAA ACAAGTCATC CGTCTGGATG TGGTTAGCTT AGTTCAAGGA 600

ACGGGGATTC GAGGACAATT TGAAGAACGC ATGCAAAAAC TCATGGAAGA AATTCGCAAA 660

CGTGAAGACA TCATCCTCTT TATCGATGAA ATCCATGAAA TTGTTGGTGC TGGTTCTGCG 720

AGTGATGGTA ATATGGACGC AGGAAATATC CTCAAGCCAG CCCTTGCTCG TGGAGAACTG 780

CAACTAGTCG GTGCTACTAC CCTCAATGAA TACCGTATCA TTGAAAAGGA TGCTGCCCTC 840

GAGCGTCGTA TGCAGCCTGT TAAAGTCGAT GAACCAACGG TGGACGAAAC AATCACTATT 900

CTCAAAGGGA TTCAAAAGAA ATACGAAGAT TACCACCACG TTCAATATAC AGATGCTGCG 960

ATTGAAGCAG CTGCAACTCT TTCCAATCGC TACATCCAAG ATCGCTTCTT GCCTGACAAG 1020

GCCATTGACC TCCTAGATGA AGCTGGTTCT AAGATGAACT TGACCTTGAA TTTTGTGGAT 1080

CCTAAAGTAA TTGATCAGCG CTTGATTGAG GCTGAAAATC TCAAGTCTCA AGCTACACGA 1140

GAAGAAGATT TTGAGAAGGC GGCCTACTTC CGCGACCAGA TTGCCAAGTA TAAGGAAATG 1200

CAAAAGAAAA AGATCACAGA CCAGGATACT CCTAGCATCA GCGAGAAAAC TATTGAGCAC 1260

ATTATCGAGC AGAAAACCAA TATCCCTGTT GGTGATTTGA AAGAGAAAGA ACAATCTCAA 1320

CTCATCCATC TAGCCGAAGA TCTCAAGTCT CATGTTATTG GTCAAGATGA TGCAGTCGAT 1380

AAGATTGCCA AGGCTATTCG CCGTAATCGT GTCGGACTTG GTACCCCTAA CCGCCCAATC 1440

GGAAGCTTCC TCTTCGTTGG GCCAACTGGT GTCGGTAAGA CAGAACTTTC CAAACAACTG 1500

GCTATCGAAC TTTTTGGTTC TGCTGATAGT ATGATTCGCT TTGATATGAG TGAATACATG 1560

GAAAAACATA GTGTGGCTAA GTTGGTCGGC GCTCCTCCAG GTTATGTTGG CTATGATGAG 1620

GCTGGTCAAT TAACTGAAAA AGTTCGCCAC AATCCATATT CTCTCATCCT TCTCGATGAA 1680

GTGGAAAAAG CTCACCCAGA TGTTATGCAC ATGTTTCTTC AAGTCTTGGA CGATGGTCGT 1740

TTGACAGACG GGCAAGGACG CACCGTTAGC TTCAAGGATG CCATCATTAT CATGACCTCA 1800

AATGCAGGTA CAGGAAAGAC CGAAGCTAGC GTTGGATTTG GTGCTGCTAG AGAAGGACGT 1860

ACCAATTCTG TCCTCGGTGA ACTCGGTAAC TTCTTTAGCC CAGAGTTTAT GAACCGTTTT 1920

GATGGCATTA TCGAATTTAA GGCTCTCAGC AAGGATAACC TCCTTCAGAT TGTCGAGCTC 1980

ATGCTAGCAG ATGTTAACAA GCGCCTCTCT AGCAACAACA TTCGTTTGGA TGTAACTGAT 2040

AAGGTCAAGG AAAAGTTGGT TGACCTAGGT TATGATCCAA AAATGGGAGC ACGCCCAcTT 2100

CGTCGGACTA TTCAAGACTA TATTGAGGAC ACAATCACTG ACTACTACCT TGAAAATCCA 2160

AGCGAAAAAG ATCTCAAAGC AGTTATGACT AGCAAGGGAA ACATTCAGAT TAAATCTGCC 2220

AAAAAAGCTG AAGTTAAAAG TTCTGAAAAA GAAAAATAAA TCCTATAAAA AAGGAGTAGA 2280

AAATGAAATT TTTCTGCTTC TTTTTTTACT AAAATAACTG TAATTTCTTG ACAGCTTGCC 2340

CTTTGTCCAT TATGATATAT AGTAGACTGA ATCTGAAATA GTACGAAACA ATTGCTAAAA 2400

CATTTATAGA AATTAATTTT ACTTTCCCAA TCGATTTGTT CTCATCTTAT TTCAATCTGC 2460

TATAGTCAAT TGAAACAAGA ACAAGACAAA AGAGCCTCAT AAAAGGTATT GCAACTTGGT 2520

AATACCTTTT TGAGGTGCTT TTTGATATGA GCCCATGTTT TCTCAATAGG ATTGTACTCA 2580

GGTGAGTAGG GAGGAAGAGG TAAAAGTTTA TACCCAAACT CTTCACACAA GAGTTCTAAC 2640

TTACCCATTC TATGGAATCT TGCATTATCC ATAATAATAA CCGATGGTGT GGTTAATGTT 2700

GGTAAGAGAA ACTTCTGAAA CCAAGCTTCA AAAAAGTCGC TCGTCATCGT CTCTTCGTAA 2760

GTCATTGGAG CGATTAACTC ACCATTCATT TGTTAGACCT GCAACCAAAG AAATTCTCTG 2820

ATATCTTCTT CCAGATACTT TGCCTCTTCT TAACTGACCT TTTAATGAGC GACCATATTC 2880

TCGATAAAAA TAAGTATCGA ATCCTGTTTC GTCAATCTAA ACAGGTGCTA GGTGCTTTAA 2940

ACTATTAAAA TTCTTAAGAA ATAAGGCTAC TTTTTCTGGG TCTTGTTCAT AGTAGGTGTA 3000

GTTCTTTTTT TTTTCGAGTG TAGCC 3025

(2) INFORMATION FOR SEQ ID NO: 169:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4104 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

TTTAAGGTTT TAAAAAAAGT TTTCGAAAGG TTTCTTCTTT ATTTTTTAAG GGAGAGATAA 60

CGTTGATATC TAAATCGTGG TCAAAGCCGG CAATTTTTCC TTTAGATGTG TATTGGTGAA 120

TATCATAATC TAAATCAGTT TTAGGACTGC TCTCCAAAAA TCCTGAGTCT GAGCCGTAGA 180

CGGAATCCAA ACAGAGGTAA ACTTGCCTGT ATCAATACTG TGTTCTTCCA TGAAGTAGAC 240

ACCAACGTAG ATGCCGATGT TTTTAGCACC CAGTGATGCT AGTTTTGCTC GAAAGTTTTC 300

GACACCTTCG TTCATATTAG ACATGGTTTT GTCTTCCACG TCAAGCCAAT AGTAACTAGG 360

GCTGTAAGGA GAGGCAGCAT TGTAGAAAAC TTCGGCAGCC TTTTCCATTT CTTGGACACT 420

TTTTCCAGCT ACATAAGCGT AGACAGCAAC TGGGACATTC CGCTTTTGAA GTTCAGTGAT 480

ATGACTCTTA TAGGCCTTGT CTATTCCATT GATAAATGAA GCATCATTTT CTTTTGTCGT 540

TTGAGCACCA CTGTGAACAC GAACAATAGC ACCTGAAATA TTTTGTGAGA GGGCATCGTA 600

GTTGATTTCC TCAGGACGCT GCCAGCCAGA GAGGTCAATA ATCGGTTTGT CTAAGTGTTT 660

CAAAGCCTGT GCTTCAATCT GTGCTATATT GGATTTTGTT TTAAACGATT GGCTGTCATT 720

AAGTGGGCGA TTGATGATTA AAATGAACAT CATAATCCCA AAAAAACTAA ATAAAATAAG 780

TGGATGAATT TGTTTTCTCA TATCTTATAA TTCTACCCTA AAAATCAAAA AAAATCAAAA 840

AAATGGGTTA AGGAAGAGAC TTTAGAGCAT TTTTTCATTC AAGAGTGCGG AATGATTTGA 900

AATATGGTAT AATAAAAGGG AATTTCTACA GAAAAGAGAA GATTATGTCA AATTTTGCCA 960

TTATTTTAGC AGCGGGTAAA GGGACTCGCA TGAAATCTGA TTTGCCAAAA GTTTTGCACA 1020

AGGTTGCGGG TATTTCTATG TTGGAACATG TTTTCCGTAG TGTGGGAGCT ATCCAACCTG 1080

AAAAGACAGT AACAGTTGTA GGACACAAGG CAGAATTGGT TGAGGAGGTC TTGGCTGGAC 1140

AGACAGAATT TGTGACTCAA TCTGAACAGT TGGGAACTGG TCATGCAGTT ATGATGACAG 1200

AGCCTATCTT AGAAGGTTTG TCAGGACACA CCTTGGTCAT TGCAGGAGAT ACTCCTTTAA 1260

TCACTGGTGA AAGCTTGAAA AACTTGATTG ATTTCCATAT CAATCATAAA AATGTGGCCA 1320

CTATCTTGAC TGCTGAAACG GATAATCCTT TTGGTTATGG ACGAATTGTT CGTAATGACA 1380

ATGCTGAGGT TCTTCGTATT GTTGAGCAGA AGGATGCTAC AGATTTTGAA AAGCAAATCA 1440

AGGAAATCAA CACTGGAACA TACGTCTTTG ACAACGAGCG TTTGTTTGAG GCTTTGAAAA 1500

ATATCAATAC CAATAACGCT CAAGGCGAAT ACTATATTAC AGACGTCATT GGTATTTTCC 1560

GTGAAACTGG TGAAAAAGTT GGCGCTTATA CTTTGAAAGA TTTTGATGAA AGTCTTGGGG 1620

TAAATGACCG TGTGGCGCTT GCGACAGCTG AGTCAGTTAT GCGTCGTCGC ATCAATCATA 1680

AACACATGGT CAACGGTGTT AGCTTTGTCA ATCCAGAAGC AACTTATATC GATATTGATG 1740

TTGAGATTGC TTCGGAAGTT CAAATCGAAG CCAATGTTAC CTTGAAAGGG CAAACGAAAA 1800

TTGGTGCTGA GACTGTTTTG ACAAACGGTA CTTATGTAGT GGACAGCACT ATCGGAGCAG 1860

GAGCGGTCAT TACCAATTCT ATGATTGAGG AAAGTAGTGT TGCAGACGGT GTGATAGTCG 1920

GTCCTTATGC TCACATTCGT CCAAATTCAA GTCTGGGTGC CCAAGTTCAT ATTGGTAACT 1980

TTGTTGAGGT GAAAGGATCT TCAATCGGTG AGAATACCAA GGCTGGTCAT TTGACTTATA 2040

TCGGAAACTG TGAAGTGGGA AGCAACGTTA ATTTCGGTGC TGGAACTATT ACAGTCAACT 2100

ATGACGGCAA AAACAAATAC AAGACAGTCA TTGGAAACAA TGTCTTTGTT GGTTCAAATT 2160

CAACCATTAT TGCACCAGTA GAACTTGGTG ACAATTCCCT CGTTGGTGCT GGTTCAACTA 2220

TTACTAAAGA CGTGCCAGCA GATGCTATTG CTATTGGTCG CGGTCGTCAG ATCAATAAAG 2280

ACGAATATGC AACACGTCTT CCTCATCATC CTAAGAACCA GTAGGAGCCT ATCATGGAGT 2340

TTGAAGAAAA AACGCTTAGC CGAAAAGAAA TCTATCAAGG ACCAATATTT AAACTGGTCC 2400

AAGATCAGGT TGAATTACCA GAAGGCAAGG GAACTGCCCA ACGGGATTTG ATTTTCCACA 2460

ATGGGGCTGT CTGTGTTTTA GCAGTAACGG ATGAACAAAA ACTTATCTTG GTCAAGCAGT 2520

ACCGCAAAGC TATCGAGGCT GTCTCTTACG AAATTCCAGC CGGAAAATTG GAAGTAGGAG 2580

AAAACACAGC CCCTGTGGCA GCTGCCCTTC GTGAATTAGA GGAAGAAACA GCCTATACAG 2640

GGAAATTAGA ACTCTTGTAC GATTTTTATT CAGCTATTGG CTTTTGTAAT GAGAAGTTAA 2700

AACTATATTT AGCAAGCGAT TTGACAAAAG TGGAAAATCC GCGTCCGCAG GATGAGGATG 2760

AAACCTTGGA AGTCCTTGAA GTGAGCTTAG AAGAAGCGAA AGAATTAATC CAATCAGGTC 2820

ATATCTGTGA TGCCAAGACA ATTATGGCTG TTCAGTATTG GGAGTTGCAG AAAAAATAGA 2880

GGAGGTCAGT ATGGGTAAAT CTTTATTAAC GGATGAAATG ATTGAAAGAG CTAATAGAGG 2940

CGAAAAAATT TCAGGTCCTC CTTTGCTAGA TGATAATGAG GAAACTAAGA TTTTACCAAC 3000

CTCTTCTTCC CGTTTTGGTT ATGCCAATCC TAAGGATCAT GGTTTTAGCC AGGAAACCTT 3060

GAAGATTCAG GTCGAACCAT CTATTCATAA AAGCCGTCGT ATTGAAAATA CCAAGAGAAA 3120

TGTCTTCAAT TCTAAGTTGA ATAAAATCTT ATTTGCGGTC ATCTTTCTCT TGATTTTGCT 3180

TGTTTTAGCA ATGAAACTTT TGTAATAGAA AAGGAATTGA AATGAAAATA GGAATTATTG 3240

CTGCTATGCC AGAAGAACTG GCTTATCTGG TCCAGCATTT AGATAATGCC CAGGAGCAAG 3300

1080

TTGTTTTTGG GAATACCTAT CATACAGGAA CCATTGCTTC TCATGAAGTC GTTCTTGTAG 3360

AAAGTGGAAT TGGTAAGGTC ATGTCTGCTA TGAGTGTGGC GATTTTGGCT GATCATTTCC 3420

AGGTGGATGC CCTTATTAAT ACGGGTTCAG CTGGGGCAGT AGCAGAAGGT ATCGCTGTTG 3480

GGGATGTCGT GATTGCTGAC AAATTAGCCT ATCATGACGT GGATGTCACA GCTTTTGGCT 3540

ATGCTTATGG ACAAATGGCG CAACAACCGC TTTATTTCGA ATCAGACAAA ACCTTTGTTG 3600

CTCAAATCCA AAAGAGTTTA TCTCAATTGG ACCAAAACTG GCATCTTGGT TTGATTGCTA 3660

CAGGAGATAG TTTTGTTGCA GGAAATGACA AGATAGAAGC GATTAAGTCC CATTTCCCAG 3720

AAGTTTTAGC CGTGGAGATG GAGGGGGCAG CTATTGCTCA AGCAGCGCAT GCCCTCAATC 3780

TCCCAGTCTT AGTCATCCGA GCTATGAGTG ACAATGCCAA CCATGAAGCA AACATCTTTT 3840

TTGATGAGTT TATTATCGAA GCTGGACGTC GCTCTGCCCA AGTCTTGTTG ACCTTTTTGA 3900

AGGCTTTAGA TTAAGCGGAA ATTTGACAGT TTTTCTAGCT TATGATAAGA TTTAAGTAAA 3960

GAAAAGCTAG AAAACGTTTC AGAGGATATT ATGAGTATTG AAATGACCGT CAGTGAGATT 4020

GCAGAGGTCT TAGGATTATC TCGCCAAGCA ATCAATAACC GTGTCAAAGA ATTACCAGAA 4080

GAAGACACAG ATAAAAATGA CAAG 4104

(2) INFORMATION FOR SEQ ID NO: 170:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8876 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

CACGGATAGG CTCGGCTTTC ATCAGTCCTC AGGCTGATTT ACTAATAGCA ACTTTCCTCG 60

ACAAAGTCCA CAGCGATACG TnTGGGTATC AATCCTACGC TTACGCTGAT ACCTTTGCTG 120

GCAGGATTGG CAACGATAGA GCTTGATTGG CTTGGAGTTA CTATTGGGCA AGGATGGTAC 180

AAACCGTAAT CCATCCACTG CTTTCAACAG TTCCTTAAAA TCCCGATCCT TGTGTTGATA 240

GCCTTTCCCT TGAAAATAGA GGTGATAATG ACAGAGTTCA TGTCGGACAA TTTTCCTAAA 300

AACGTCCAAC CCCAGTTCCT GATAAACCTT GGGATTAAAA TCCAAATGCC CATCTTTGGG 360

GAAAAATCGC CCACCTGTCG AACGTAGACG CCTATTCCAC TGGACATGAT GGATAAAAGG 420

TCTGCCGAAG TCTTCTAGTG AAACCTGCTT GACGTAATCA GTCAGTTTCA TTTGGAGCTA 480

GGAGAGACAG ATTAACTTTT TCACGTTCAG TATCAATTTT CTTAACCCAA ACGCTCACCA 540

AATCTCCAAC TGCCACCACT TGACTAGGGT GTTTGATAAA CTTGCGACTC ATATGGGAAA 600

TATGGATGAG ACCGTCCTCA TGAATTCCGA TATCAACAAA AGCACCGAAA TCAACAACGT 660

TACGCACCAC TCCTTCTAGC TTTTGTCCAA CCACTAAGTC CTTGATATCT AGGACATCTT 720

GGCGAaCACA GGTGCGTCAA AGGAATCACG GAAATCTCGA CCTGGTTTGA GAAGATCTGC 780

AATGATATCT TTAAGAGTTT CTGGACCAAG GTCTAACTCT TGCGCCATTT CCTTGACTGA 840

AAGCGACTTG AGTTTGCTTT GGGCTTCTTC GTTTAGGTCT TTAATATCTA AACGTTTGAA 900

GAGTTCCTTA ACTGCAGTGT AATTCTCTGG GTGAACTCCT GTATTATCAA GGATATTGCT 960

ACTTTCAGGG ATACGAAGGA AACCAGCAGC CTGCTCAAAG GCCTTGGCTC CCAGACGAGG 1020

AACTTTCTTG ATTTGGGCGC GTGAAGTGAT TTTTCCTTCT TCCTCGCGGT ATTTGACAAT 1080

ATTTTCAGAG ATAGTTTTGT TGAGTCCAGC TACGTGTGAA AGAAGAGCTG GGCTAGCTGT 1140

ATTGACATTG ACACCAACTT GGTTAACCAC TGTATCGACA ACAAAGTCCA GACTCTCAGA 1200

TAGTTTCTTC TGACTGACAT CGTGTTGGTA TTGACCGACA CCAATTGACT TAGGATCGAT 1260

TTTGACCAAT TCCGCAAGAG GATCTTGCAA ACGACGGGCG ATAGAAATGG CAGAGCGTTT 1320

TTCAACGGTC AAGTCTGGAA ACTCCTGACG AGCAAGTTCG CTGGCAGAAT AGACAGAAGC 1380

ACCACTTTCA TTAACGATAA CATAGCTGAC TTCAGGGAAA TCTTTCAGAA CTTCCGCTAC 1440

AAAAGCTTCA CTTTCACGAC TGGCCGTTCC ATTTCCAATG GCAATAATCT CTACACCGTA 1500

TTGACCAATT AAATCTGCTA AATCTTTCTT GGCTTCTTCG ATTTGACGAG CTGATGCTGG 1560

TTTAACAGGA TAAATAACCT GAGTTGTCAG CATTTTTCCT GTTGCATCCA CGACAGCTAG 1620

CTTGGCACCT GTACGAAAGG CTGGGTCAAA TCCAAGAACC ACGCGCCCTT TCAGTGGAGC 1680

AACCAAGAGG AGATTGCGCA GATTGTCAGA AAAAAGTTGG ATAGCTCCTT CTTCAGCTTT 1740

CTCAGTTAAT TCTGTCCGAA TACGACGCTC GATAGCAGGC AAGACCTTTT TCTTAACGGA 1800

TTGCTGAACA ACTTCATCAA TATAAGCATT TTTCACCTTG AAACGAGTAG CAAAGAAGGC 1860

AAGAATACGG TCCGTCGCAT GTTCAAAACC GATCTTCAAG ACACCAAGTT TCTCCCCACG 1920

ATTGAGAGCC AAGGTACGAT AGCCTTGCAT AGTTCCAACT GTCTCTGAAA AATCATAATA 1980

AATCTGAAAA ACCTGCTTTT CATCAAGACT TTCATCCTTG GCTTGAGAAG TAAGTTTAGA 2040

GTGTCTCAGC ACTTCCTGAT AAGTCATAGA ACGCAAGGTC ACATCTTCCG ATAAGGCTTC 2100

GACCAAAATA TCAACTGCAC CGGTCAAGGC TTCCTTGCCA GTCGCAAATC CTTCACAGAC 2160

GAACTTTTCA GCTTCTTTCT CTAAGTCAAC TATATTCTGC AAAATCAAGC GAGCAAGAGG 2220

AAAGAGTCCA GCTTCACGGG CAATGGTTGC CTTGGTACGA CGCTTTTCCT TATAAGGAAG 2280

ATAGAGTTCT TCAACGTCTG CTAATTTTTC GGCAACTAAG ATAGCTTCTT CCAATTCCTT 2340

GGTCAACTTA CCTTGTTCTT GAATCTTAGC TAAGACAGCT TCCTTACGGT CATTGAGATT 2400

TGTCAGACTT TTATCCAAAT CAATAATAGC CTTAATCGCC ACCTCATCCA GACTACCAGT 2460

CATGTCCTTG CGATAACGCG CGATAAAGGG AATAGTCGCC CCTTCAGCTG TCAAACTTAG 2520

AACGGTATCA ATTTGCTTTA ACGTCACTCC CAAATCCTGA GAGATTTTTT CATATTTTTT 2580

ATCCATAAAT CTATTATACC ACAAGCTAAA CGTTTCAAAT TAACTCGTAG AACATTTAAA 2640

AAATATGTAG GAAATAGATT TATATGCTAC AGCGCAATAA CTTGCACTTA AAGAGCATTG 2700

CCACCTTTTT TTAACCAAGC CATGATATCA AAAGTATTTA ATGGATCAGA CATAATAGCC 2760

AGTTCTGGAA GATGTTCCTG ACCTGGAATA ACACATTGAC TTTTCAAATT TTTATATGGA 2820

CGATTGACTA AAATTAATTT ATTAGAATAA GGAAGATTAT CCATCTTATT TAAAATTTCT 2880

TCACTAGCTG AATCTTTATT ATCAAATTTA AAATAAAGAT TATTCCAATT TATGCGTTTT 2940

TTTCTTTTTT CCCACTTAGT TCGTGCTTCT TCAATACTAG AATAATGTAG AAAATGAATA 3000

TCTATATCTC CTAAGTGCCC CAAAGGATAA ACTTCATGAG TCCAGCTCGG TGAAATAAGT 3060

TCCTCTTCGA AAACAAGTTC TTGTTCCATA TAATAACGAA AATGCTTTGT AAGTTTATAA 3120

TAATCATCAG GAAGAATAAA TAAACCAACA AAAGGTGTTC TATATTGAAA ACCAAGCTGT 3180

TTATAAATTA ATCCTCCAAC ACAATTATTA CTTATAATCG TAAAATCTAA TCTATCAAGC 3240

TCAAGAAAAG GGAAAATTCC TTTCTCTGCA GCTATTAACT TATGATAAAC AATATCAGAA 3300

TCTAAATATT CACCGTCATT TTTTAACCAA GCACTAAAAT TTGCCAATTC TTGAATATAT 3360

TGTTTTTTCG CTCTTTCTAT ATCATAGTTT TCTAAGACGG CGCAATCTTT GATTCTATTT 3420

TCATAATTTT CTAATATGAT TTTGTAGGAG TCTTTTAGAG GTTTAGCATC TATAACAGGT 3480

TTATAGATAT ATGTCGGGAA ATTAATATAG GTTGCAGTTT TAGAGTGAAT ATAAAGTCTC 3540

CAAATAAGGT TGTTTATATC AAATTGATTT ATTTTTCGTA AAAGCTTACT ATTGAATAAT 3600

TTTCCAAATA ATGAGCGATA TTGTTTTCTA ATTCGATGAT CTGTATCATC CATCTTTTGT 3660

AAAACTTGAA CATTCGTTAA ATTTTCTGTC AACCAATTAT CCCCCCAAAA AGGATAAAAG 3720

TAAAATACTC CATCAACCAA ATCAGCAAAA TGACCAAGAA CAACATCAGA ATCGGATAAT 3780

TTTATCGCAT GATACATCTT TTCAAATGTC CAATCAAATA ATGAATCATT TGAAGATAGA 3840

AACGTAATAT AATCTCCTGT AATCATATCA GACAACTCAG CAAAAGAATT CTCATCTATA 3900

ATCTTAATAT TAAATGATAG ATTCATCTGT TGGCTAATGG AAGCTATCTC CTCTGTAGAT 3960

TGATTTACAA TAATAACTTC TATATCTTTT AATGTTTGTC TCTCCACTAT TGACAAAGAC 4020

TCTAATAAAC TATTTTTATC TCCTTGATGT AACAAAACAA CACTAATTGA GTAAGTCAGT 4080

TTGACTACCT CCCATAATTT TCTGATAATG ATTTTCTTTT TATTTAATTA TAGCACAATT 4140

ATGATATATA TCAGGTAATA TCAAGCTATA TTATCTCTTA GCTACTCAAT TTGAAATTTT 4200

AACTTTTCCC TTTTCCGCAA AATAATAGTA TAATAGAGGT AGAATCTAGA ATCGAGGTAC 4260

ACCTATGGCT GTCAAATTTA CAAAACGAGA CGACTTGGAC AAGATGTTTG AAGAGTTTGC 4320

TAAACTCCCT GATTTGAAAC AAGTTACTTT CCCTGATGAC AAAGAGAAAA AAGTCAAAGC 4380

AGAAAAGAAA AACTAGATGA CTGCTTTTCA ACAACTCCCA TCTAGTGTAC TTCAAACTGG 4440

AGCCATTTTT CTCTCCATTA TCATTGAAGC CCTTCCCTTC GTTCTGATAG GAAGCATTGT 4500

CTCAGGGCTG ATTGAAGTTT ATATCACACC TGACAAGGTT TATCATTTTC TCCCTCGAAA 4560

TCGTTGGGGG AGAATCTTTT TTGGGACCTT TGTCGGTATA CTTTTCCCTT CTTGTGAATG 4620

TGGAATCGTC CCCATCATCA ATCGTTTTCT GGAAAAAAAG GTTCCAAGTT ACACGGCCGT 4680

TCCTTTTCTT GTGACAGCAC CTGTTATCAA TCCCATTGTT CTTTTTGCGA CCTATTCTGC 4740

CTTTGGCAAC TCCTTCCATG TCGCCCTATT ACGAGCTCTG GGTTCCATTC TTGTGGCTGT 4800

AATACTAGGA ATTTTTCTAG GATTTTTCTG GCAAGAACCG ATTCAGAAAG AAAATCGTCT 4860

GGCTTGTCAT GAGCATGATT TTTCTTACTT GAGTTCTGCA AAAAAAGTTT TTCAAGTCTT 4920

TGTGCAGGCC ATTGATGAAT TTTTTGATAC GGGGCGTTAT TTGGTATTTG GCTGCCTCTT 4980

TGCTTCTATA ATACAGGTCT ACGTTCCGAC TCGGATTCTG ACCTCTATCA GTGCGACCCC 5040

TCTTTTTGCC ATCCTGCTCT TGATGATTTT AGCCTTTCTT CTTTCGCTCT GTAGTGAGGC 5100

GGATGCCTTT ATAGGTGCTT CTCTTCTCTC GAGTTTCGGT TTGGCACCAG TTCTGGCCTT 5160

TCTCGTCATT GGTCCAATGC TGGATATCAA AAATATTCTC ATGATGAAAA ATTACTTGAA 5220

AGCACGATTT ATCAGTCACT TCATAACAAT TGTAACTCTT GTCGTCTTAG TCTATTCTCT 5280

CTTGATTGGA GTTATCCTAT GATTCGATTT TTAGTTTTAG CTGGCTATTT TGAACTGACT 5340

ATTTACCTCC ATCTGTCGGG CAAACTAAAC CAGTACATCA ACATGCACTA TTCCTATCTG 5400

GCCTATATCT CCATGGTGCT TTCTTTTATC TTGGCTATCG TTCAATTGTA TATCTGGATG 5460

AAGCAAGTCA AAACCCACAG TCATCTGAAC AGCCGATTAG CCAAGATAAC GAGTATTTCT 5520

CTTCTGGCTA TTCCACTTGT CATCGGCTTA ACTTTCCCAA CTGTTAGCTT GGATTCTCAG 5580

ACTGTTTCTG CTAAAGGTTA TCATTTCCCC CTATCGGAAG GAACGGATCT AGCCATTCAG 5640

ACAAGCGAAG GGACGACAAG CCAATATTTG AAACCAGATA CCAGTTCTTA TTTTTCAAAA 5700

TCAGCCTATG AAAAGGAAAT GCGAACGGCG GCGGATAAAT ACTTATCCCA AGATAGTATT 5760

CAGATCACTA ATGAAAACTA TATGGAAGTC ATGGAGGCTA TCTACGACTA TCCAGATGAG 5820

TTTGAGGGCA AGACAATCCA GTTTACAGGC TTTGTCTATA ACGACCCCAG TCATGCCAAT 5880

AGTCAATTTC TGTTCCGATT CGGCATTATC CACTGTATCG CAGATTCTGG TGTCTATGGA 5940

TTGCTGACCA AGGGCAATAC CCGGCAGTAT GAAAACAACA CTTGGATAAC AGCCAAAGGA 6000

AAACTGGTCA ATCACTACCA TAAAGAACTC AAACAAAACC TTCCAACCTT GGAAATCGAC 6060

AGCTTTACCA AAGTCGATAA ACCAGAAAAT CCCTATGTAT ATAGAGCTTT TTAAGAAAAT 6120

CAAGATAAAA ACGAACAAGT TCTCTTCTGA ATAACAGAAA AAGAGCCTGT TCGTTTTTTG 6180

TTATATGAAA ATTAGTGACT TGTAGATTTT CATCTTATAC CATTCCCAGC AATACAAGTA 6240

GCTCATAGAA AATAAGCGAG CCACTCATTC ATTAGACTAG CGATTTCTTT AGGTGCTTGA 6300

GTATAAAGCT CATGGCCAAA GTTTTCTAAA AAAATAGTAT CAAAATAGTC TGGCAATTCT 6360

TTTAGGGCTT CCTCTCTCCA TGTAGCTTCA TTAGGATAGC GAGGACTAAT AAACAAGGTA 6420

TCTCCCACTT CTCTCTTAAA AGCTTGTATT TTTCTCCGTA GcGGAGTATC GCTTCTATAT 6480

TTTCATAATT TATAGCCAAC TCATATCTAT TATACTCAAC ATTCCAGTGA TAAGACTGTC 6540

TTACAGCTTT CTCCATATTT TCTGACCAAT GCTTTGCTTC AGATTTTTCT TTAGAAGTAA 6600

GAACATCTAA GTCCGAAACA ATTTGAGATT TGATATAATT TTTAGTTTCC TCTAACTCTG 6660

TATCCAAAGG TAAAATCTTA TCTAAATCTA GATAGCCACC ATCCAAAAGA ATCAGTTTCT 6720

TTACTTCTTC AAATTCCGAT GCGAAATAAC GAGCTAAATC TCCTCCAAGA GAATGGCCTA 6780

TCAGACAGAT AGATTCTTCC TCTACAATTT CATTTTTAAA CCATGATTTC AATTCTGTTT 6840

CATCTCGAAG ATGCTTTTCA TATGGATTTA GAAAATAGAC CTGCGAATCT AGTTCTTGAA 6900

GAAAATCCTT GCTATGATAG GCATTGCTTC CCAAACCGCC AATAAAATAT TTTTTCATTC 6960

TCTACTTAAT ACTATGCTTA TTCATCTTTT GTTCAAAGAT AGTTGTGATA ATCTGACGCA 7020

ATTCTTCGCG TTTTTTTTCT GGAATCTCAC CACTTGTTTG AGCTACAGCG TAGAGTTCAG 7080

GGTATTCAAT TGAAATGCGT TTAATCGTAC GTGTTGTAGC ATGTTTTCTG ACAAAAAACG 7140

GGATTCGCTT AATCAAGTCT TGTGGGACTA GCGCCAGAAT CTTCTCAGTA GTTTCTTTGT 7200

CACTAATATT AGACATTGTA AGCCTTTTCT TAATCATTTC CTGTTCTTTT TCTGTAAAAT 7260

CTTTTAATTC CATTCGATTA GTCCTCCTAT TTTCTCTAAG TTAAATTATG TACTAATACA 7320

GATGAAACTA CAAAGAATAA ACTTTAAGAA ATCTTCTCAC TGATAAGATT TTAGCATTAG 7380

ACTTCCTGCG AAACAAAATA TGGTATAGTA GTTCTATGAA TTATGAAGCA AGTAAACAAC 7440

TAACTGATGC ACGATTTAAA CGTCTTGTTG GTGTTCAGCG CACGACTTTT GAAGAGATAT 7500

TAGCTGTATT AAAAACAGCT TATCAACTTA AACACGCAAA AGGTGGACGA AAACCTAAAT 7560

TAAGCCTAGA AGACCTTCTT ATGGCCACTC TTCAATATGT GCGAGAATAC CGCACTTATG 7620

AAGAAATTGC GGCTGATTTT GGTATTCACG AAAGCAACTT AATCCGTCGG AGCCAATGGG 7680

TTTAAGTAAC TCTTGTTCAA AGTGGTGTTA CGATTTCAAG AACTCCTCTC AGTTCTGAGG 7740

ACACGGTAAT GATTGATAGC CATTCCCATC AATATCGTAT CTTTGGACAT AGCCAATAAA 7800

TGTTTCATTT TTGCGTGGTT TCTGGCTATT AACGATTGAA ATAACCCACC AACTTATCAA 7860

AAATAGAAAT AAAAATCCTA AGATTACTGT CATATCATAA CACTATTAAA GTTTAACCCA 7920

CTTATCATTA TCCATGATAA AAGGCTTAGC CAGTCCCTCG CCTGTATAAT CCGCATACTT 7980

GGTGCCCAAA TACTTGTAGC AATCTTCCTT ACTAGCAAAT TTAATCGCTT GGTAGGGCTC 8040

TTCGAAAGTC AATTTCTCTA CAAATAAGAA ACCGTCATCA GCAGGTACTA AGACCCCAAC 8100

GTGGCCTACA AACAGATACT CGCCATCCAA ATTGTCGTGC AAGACTACAG ACAGCATTCG 8160

AGCTTTTTCA TTGAATTGAA ATTGTGAGAA GAATGCTTCC ATCTTTTCAG CGTGAACCTT 8220

GACATCTGTA GTTGACTCAG TTGGAACTCT CGAAAATAGA ATATCAAACT CTTCCTTATC 8280

TTGTGAATCA AAGACCTTTC CTTTATCAAT CGCATCATTA TCTAGGAAAA GCAACTGGTC 8340

ATTCTTTTCA AGCTTTGGAA TGGTGACTGA ATTTTTCAAA AGACAATAAC TATTGATACG 8400

GCAGTTGGTC CCAACAAAAT CGCCCTTCTT TTGATTCCAG AGATGACTGA TTTTCTCAAC 8460

ATCGTATTCG GTGTGAGTAA AGGAAGTGAA ATCTCCTGAT AAGCCAGTTG AGCCGACAAT 8520

GGTATTATAG TCATTAACGA GATTAAAAAA TGCATCAACA CTATTTGGAT CCAAGTGAGC 8580

TGATAAGAGA GATTTGACCT CTTCTGTACT TACCTGGTTG TTTAGGTTGG TGTATGAAGC 8640

TTTCCATGGA ACTTTCGCTG AACTGCTTTG CCTTTGATTC GTCCCCTCAG AAGTAGCATG 8700

TTGTTGTTGA CAAGCAGCCA AGCCTAAAAA CAAGGCTGAA CAGATTCCTA ATGTGGCTAA 8760

TTTTCTTGAT TTCTTCATTT CTTTCTCCTA AATGTCTTGG ATTAAAGTTT CTTTAACTAT 8820

TGCTTTACAG ATATTGATTA CTTTCTCATT TAATGTGTTC ATCGTCTTTC CTCCGG 8876

(2) INFORMATION FOR SEQ ID NO: 171:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14736 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

CGCAAACTTT CGCGGTCGGA AGGTAGTTTT ATGACACGAT TTGAGATACG AGATGATTTC 60

TATCTCGATG GAAAATCATT TAAGATTTTA TCTGGTGCCA TTCATTATTT TAGGGTTCCT 120

CCAGAGGATT GGTATCATTC GCTCTATAAC TTGAAGGCTC TTGGTTTTAA TACGGTAGAG 180

1086

ACTTATGTTG CTTGGAATTT ACACGAGCCT TGTGAAGGTG AGTTTCATTT TGAAGGTGAT 240

CTGGATTTAG AGAAAATTTCT CCAAATAGCG CAGGATTTGG GTCTCTACGC AATTGTGCGT 300

CCGTCTCCAT TTATCTGTGC GGAATGGGAA TTCGGTGGCT TACCAGCTTG GCTCTTGACC 360

AAGAACATGC GAATTCGCTC ATCCGACCCA GCATATATCG AGGCAGTTGG TCGCTACTAT 420

GATCAGTTAT TGCCAAGACT GGTGCCTCGT TTGTTGGACA ATGGTGGCAA TATTCTCATG 480

ATGCAGGTTG AAAATGAGTA TGGTTCTTAC GGAGAAGATA AGGCTTACCT GAGAGCGATT 540

CGACAGCTAA TGGAAGAGTG TGGCGTAACC TGTCCCCTCT TTACATCAGA TGGTCCATGG 600

CGAGCTACTC TGAAAGCTGG AACCTTAATT GAAGAGGACC TCTTTGTAAC AGGAAACTTT 660

GGTTCTAAGG CACCTTACAA CTTTTCGCAG ATGCAGGAAT TCTTTGATGA ACATGGTAAG 720

AAATGGCCAC TCATGTGTAT GGAGTTCTGG GATGGTTGGT TCAATCGCTG GAAAGAACCG 780

ATTATCACAC GGGATCCTAA GGAATTGGCA GATGCAGTTC GAGAGGTTTT GGAACAAGGC 840

TCTATCAATC TTTACATGTT CCACGGTGGT ACAAACTTTG GTTTCATGAA TGGTTGCTCA 900

GCTCGAGGAA CTTTGGACCT GCCACAAGTT ACGTCTTATG ATTACGATGC CCTTCTGGAT 960

GAAGAAGGAA ATCCAACTGC TAAATATCTT GCAGTCAAGA AGATGATGGC AACACATTTT 1020

TCAGAGTATC CGCAGTTGGA ACCACTCTAC AAAGAGAGTA TGGAGTTGGA TGCTATTCCA 1080

CTAGTTGAAA AAGTTTCTTT GTTTGAAACC TTAGATAGCT TGTCAAGTCC TGTAGAAAGT 1140

CTCTATCCTC AAAAGATGGA GGAGCTGGGA CAAAGTTATG GCTACCTACT TTATCGAACA 1200

GAAACAAACT GGGATGCAGA AGAAGAAAGA CTTCGTATCA TTGATGGTCG AGATAGGGCC 1260

CAGCTGTATG TCGATGGTCA GTGGGTTAAA ACTCAATATC AGACAGAGAT TGGGGAAGAT 1320

ATTTTTTATC AAGGTAAAAA GAAAGGGCTA TCTAGGTTAG ATATCTTGAT AGAAAATATG 1380

GGGCGTGTCA ACTATGGGCA TAAGTTCTTA GCGGATACGC AACGTAAGGG AATTCGGACA 1440

GGGGTCTGTA AGGATCTGCA TTTCTTACTA AACTGGAAAC ACTATCCACT CCCACTAGAC 1500

AATCCTGAGA AAATTGATTT TTCAAAAGGA TGGACTCAAG GACAACCAGC CTTTTACGCT 1560

TATGACTTTA CAGTCGAAGA GCCAAAAGAT ACTTACCTAG ACTTGTCTGA GTTTGGTAAG 1620

GGGGTTGCCT TTGTCAATGG GCAGAATCTA GGACGTTTTT GGAACGTTGG CCCAACTCTC 1680

TCACTTTATA TCCCTCATAG CTATCTCAAG GAAGGTGCCA ACCGCATCAT TATCTTTGAA 1740

ACAGAAGGTC AATATAAAGA AGAGATTCAT TTAACTCGTA AACCTACACT AAAACATATA 1800

AAGGGGGAAA ACTTATGACA ATTGTAGGAT GCCGTATTGA TGGACGTTTG ATCCACGGAC 1860

AAGTAGCCAA TCTTTGGGCT GGAAAACTAA ATGTTTCACG CATTATGGTT GTAGACGACG 1920

AAGTTGTCAA CAACGATATT GAAAAGAGTG GTTTGAAACT TGCGACACCA CCAGGTGTGA 1980

AATTGAGTAT TTTGCCAGTT GAGAAAGCTG CAGCCAATAT TCTTGGTGGC AAATACGATA 2040

GCCAACGTCT CTTTATCGTG GCTCGTAAAC CAGACCGCTT CCTTGGTTTG GTAGAAGCAG 2100

GTGTACCACT TGAAACCCTT AATGTTGGGA ATATGTCTCA AACACCAGAA ACTCGTTCTA 2160

TTACACGTTC TATCAACGTA GTAGACAAGG ATGTGGAAGA CTTCCACAAA CTGGCAGAAA 2220

AAGGTGTTAA ACTTACTGCT CAGATGGTTC CAAATGATCC AATTTCAGAC TTTTTGAGCT 2280

TATTAAAATA GGAAAAAAAT TTTTAGGAGG TCATTGTTAT GATACAATGG TGGCAAATTT 2340

TACTTCTCAC TTTGTACTCA GCTTATCAAA TCTGTGATGA GTTGACGATC GTTTCATCTG 2400

CAGGTTCCCC TGTATTTGCT GGTTTCATTA CTGGTTTAAT CATGGGAGAT GTGACTACTG 2460

GTTTACTTAT CGGTGGTAAC TTGCAACTGT TCGTTCTTGG GGTTGGTACC TTCGGTGGTG 2520

CTTCTCGTAT CGACGCAACT TCTGGTGCGG TTCTTGCGAC ACCTTCTCTG TTTCACAAGG 2580

AATTGATGCA CCGCTTGCCA TTACTACAAT CGCTGTACCA GTAGCAGCTC TCTTGACTTA 2640

CTTCGACGTT CTTGGTCGTA TGACTACTAC CTTCTTCGCT CACCGTGTGG ATGCTGCAAT 2700

CGAACGCTTT GACTATAAAG GTATTGAACG CAACTACTTG CTTGGTGCGA TTCCGTGGGC 2760

TCTATCTCGT GCCCTTCCAG TCTTCTTTGC CCTTGCTTTT GGTGGTGCCT TTGTACAATC 2820

AGTAGTAGAC TTCGTTGAAG CCTACAAATG GGTTGCAGAT GGCTTGACAC TTGCAGGACG 2880

TATGCTTCCA GGTCTTGGAT TTGCAATCTT GCTTCGTTAC CTTCCAGTTA AACGTAACCT 2940

TCACTACCTT GCTATGGGAT TTGGTTTGAC AGCTATGTTG ACTGTTCTTT ACTCATATGT 3000

AACAGGTCTT GGTGGCGCTG TTGCTGGTAT CGTAGGTACT CTTCCTGCTG AAGTTGCTGA 3060

AAAAATTGGT TTCGTGAACA ACTTCAAAGG TTTGTCTATG ATTGGTATTT CTATCGTAGG 3120

TATTTTCCTT GCAGTGCTTC ACTTCAAAAA TAGCCAAAAA GTAGCTGTAG CAGCACCTTC 3180

TACACCATCA GAAAGTGGGG AAATCGAAGA TGACGAATTC TAATTACAAA CTTACAAAAG 3240

AAGATTTTAA TCAAATCAAC AAACGTAGCT TGTTTACTTT CCAATTAGGT TGGAACTACG 3300

AACGTATGCA AGCTTCTGGT TACCTTTACA TGATCTTGCC TCAGTTGCGT AAAATGTATG 3360

GTGATGGAAC TCCTGAATTG AAAGAAATGA TGAAAGTTCA TACTCAATTC TTCAATACTT 3420

CACCATTCTT CCATACCATT ATCGCTGGTT TTGACCTTGC CATGGAAGAA AAAGATGGTG 3480

TAGGTTCAAA AGACGCCGTT AACGGTATCA AGACAGGTTT GATGGGACCA TTCGCTCCTC 3540

TTGGGGATAC AATCTTTGGT TCACTTGTAC CTGCTATCAT GGGGTCAGTC GCAGCAACTA 3600

TGGCTATCGC TGGCCAACCT TGGGGGATCT TCCTTTGGAT TGCAGTTGCA GTAGCGTATG 3660

ACATCTTCCG TTGGAAACAG TTGGAATTTG CTTACAAAGA AGGGGTTAAC CTTATCAACA 3720

ACATGCAAAG TACCTTGACA GCTTTGATTG ACGCTGCATC TGTACTTGGT GTCTTCATGA 3780

TGGGTGCTCT TGTAGCAACA GTGATTAACT TTGAAATTTC TTACAAGTTG CCAATCGGTG 3840

AAAAGATGAT TGATTTCCAA GACATCTTGA ACCAAATCTT CCCACGTTTG CTTCCAGCAA 3900

TCTTTACTGC CTTTATCTTC TGGTTGCTTG GTAAGAAAGG TATGAACTCT ACTAAAGCTA 3960

TCGGTATTAT TATCGTACTT GCTTTGGCTC TTTCTGCCCT TGGTCACTTT GCACTTGGAA 4020

TGTAATTCCT TATGACTAAA TCATTAATTT TGGTGAGCCA TGGTCGCTTC TGTGAGGAGC 4080

TTAGAGGTAG CACAGAAATG ATTATGGGCC CACAAGACAA CATTTACACA GTAGCTCTTC 4140

TTCCAGAAGA TGGCCCAGAA GAATTTACTG CTAAATTTGA AGCTGTTATT GAAGGATTGG 4200

ATGATTTCCT AGTCTTTGCG GATCTTCTCG GTGGGACACC TTGTAATGTG GTGAGTCGCT 4260

TGATCATGGA AGGTCGTGAT ATTGACCTTT ACGCAGGGAT GAATCTTCCA ATGGTGATTG 4320

AATTTATCAA TGCGAGCCTT ACAGGCGCAG ATGCGGACTA CAAGAGCCGT GCTGCAGAAA 4380

GCATTGTGAA AGTTAATGAC CTGTTAGCGG GCTTCGATGA TGACGAAGAT GAATAATACT 4440

CTTCGAAAAT CTCTTCAAAC TACGTCAACG TCGCCTTGCC GTAGgTATAT GTTACTGACT 4500

TCGTCAGTCT TATCCGGCAA CCTCAAAACG GTGTTTTGAG CTGACTTCGT CAGTCTTATC 4560

CGGCAACCTC AAAGCAGTGC TTTGAGCAGC CTGCGGCTAG TTTCCTACAG ATTTTAGTTG 4620

GAACTCGATT CAATTCATGT GACAACGTGA AAATCGTTAG AGCATTTTAT ATAGAATATA 4680

CATGGGAATG TAGCTTACTC CCATTCCCAT ATTTAATAGA AAAAGAGGAA CTCAATGCTA 4740

CATTATACAA AAGAAGACTT GCTCGAATTG GGTGCAGAAA TCACTACGCG TGAAATCTAC 4800

CAACAGCCTG ATGTATGGAG AGAAGCTTTT GAATTTTATC AAGCAAAACG TGAAGAAATT 4860

GCAGCCTTCC TACAAGAAAT CGCTGATAAA CATGACTATA TTAAGGTTAT CTTGACAGGT 4920

GCTGGGACTT CTGCTTATGT GGGAGATACC TTGCTACCTT ATTTTAAGGA AGTCTATGAC 4980

GAACGCAAAT GGAATTTCAA TGCTATTGCG ACAACAGATA TCGTTGCCAA TCCAGCAACC 5040

TATTTGAAAA AAGATGTGGC AACTGTCCTT GTGTCTTTTG CTCGTAGTGG GAATTCGCCT 5100

GAAAGTTTGG CGACTGTTGA TTTGGCCAAA TCCTTGGTGG ATGAGCTTTA TCAAGTGACG 5160

ATTACTTGTG CAGCAGATGG TAAATTGGCT CTTCAAGCTC ACGGTGATGA TCGTAATCTC 5220

TTGCTCTTGC AACCAGCTGT CTCTAATGAT GCTGGATTTG CCATGACTTC TAGCTTTACG 5280

TCTATGATGT TGACAACTCT CTTGGTCTTT GATCCTACAG AATTTGCTGT TAAGTCTGAA 5340

CGTTTTGAAG TTGTATCTAG TCTTGCCCGT AAAGTTTTAG ACAAGGCAGA AGATGTCAAA 5400

GAGCTCGTTG ATTTAGACTT TAACCGTGTC ATCTATCTAG GCGCTGGTCC TTTCTTTGGA 5460

CTTGCTCATG AAGCTCAGCT CAAGATTTTG GAATTAACTG CTGGTCAAGT TGCGACCATG 5520

TATGAAAGCC CAGTTGGCTT CCGTCACGGT CCAAAATCTC TTATCAACGA CAATACAGTT 5580

GTTTTGGTCT TTGGTACAAC GACAGACTAC ACTCGTAAGT ACGACTTGGA CTTGGTTCGT 5640

GAAGTTGCTG GTGACCAGAT TGCTCGTCGT GTTGTGCTTT TGAGTGATCA AGCTTTTGGT 5700

CTTGAAAATG TCAAAGAAGT GGCCCTTGGT TGTGGCGGTG TCTTGAATGA TATTTACCGT 5760

GTCTTCCCTT ACATCGTTTA TGCCCAACTC TTTGCTTTAT TGACTTCACT CAAGGTAGAA 5820

AATAAACCAG ATACACCGTC TCCTACAGGT ACAGTAAACC GTGTAGTACA AGGTGTCATA 5880

ATTCACGAAT ATCAAAAGTA AGACAGTGTT TATGAATTCT TGACAAGAGG ATTTGTAAAT 5940

TATCAGATAA ACCATAGATT GTCAGTACGC TTTCTATGGT TTGTTTGCTT GAGAGAAATA 6000

GTAAAAGGAG AACAGAATGA AAGCATACAC AGAGCGTGTA TTTGGAAATG TTGAGGGTGA 6060

GGATGTCTTG GCCTATCGAT TTGAGACAGA CGGTGGCTAC CAACTTGAGG TTATGACTTA 6120

TGGTGCGACT ATCTTGCGCT ATGTCGCACC TGACAAGGCT GGAAATTTTG CCAATGTTAT 6180

CTTGGGATTT GATGACTTTG ATAGTTATGT AGGCAATAGT CCCAAGCATG GAGCAAGTGT 6240

AGGTCCTGTA GCGGGTCGTA TTGCAGGTGC GACCTTTGAG CTCAATGGTA AGACCTATGA 6300

CCTTGAGGTT AATAATGCTA GCAACTGTAA TCACAGTGGT TCAACTGGTT GGGATTCCAG 6360

CTTGTTTGAA GTTGAAGAAG TAAGCGATCA TGGCTTGACT CTCTACACAG AGCGTACAGA 6420

TGGGACAGGA GGGTTCCCTG GAAATCTCAA GATTTGGATC AGTTATCACT TGGAAGAAAC 6480

TGGTGCCTAT GAAATCAGCT ACAAGGTAAC GACCGATCAG GATACGCTGG TCAATCCAAC 6540

CAACCACAGC TATTTCAACT TGTCTGGTGA TTTCACGCAG ACGATTGACC GTCATGTCTT 6600

CCAACTAAAC ACAGAGGGCA TTTACTCAAT CGCTCCTGAC GGTGTTCCTG CCAAAACTCC 6660

AGAAGCCAAC CGTGATGTGG TCAAACACGT CTACAATGGT ACCTTGTTGA AGGATATCTT 6720

TGCAGAAGAA GATGAGCAAA TCCAGCTGGC ATCAGGTTTG GATCATCCAT TTGCCCTTCC 6780

TGCAGGCCAT GACAATGCTG GATTCCTTTA TGACCAAAAT TCAGGTCGCT TCCTGCTTTT 6840

CAAGACAGAA GCTCCTTGCT TTGTGGTCTA CACAGCAAAC TTTGTGGATG AAAGTGTCAT 6900

CATAGGAGGT CAGCCAATGC TACAGCACAA TGGGATTGCT CTTGAAGCGC AAGCTTTACC 6960

AGATGCCATT CACAGTGACC TTAAAGGCCA AGTCATTCTT AAAGCTGGTC AAACCTTCAC 7020

CAGTAAGACA CGTTATGAAC TTGTTGTGAA GTAAAAGAGT CATTGCGCCT ACTTTTGGGA 7080

GCTAGGAATA GGTACGCAGA GACAAATAGT AGGAAAATAT GATATAACTA AGCGTTGAAA 7140

GCTATCTGTT AATATAATAT TCAAACTACA ATAAGGAGTA AGAAAGAAAC GAAGAAAATT 7200

GTATTTGCTA GTGCCTTGGC TTTGACCTTG GCTGGAGCAG TTTTGACAAA TGATGTTTTT 7260

1090

GCGAACGACA GACTTGTGGC AACACAAACT ACTGATGGTA AAAATGAAAA TGTATTGACC 7320

TCAGAGGTGC TAAAACCTTC TAGTGGCAAT GTTTTGGTTG GAATCAAAGG AGAATTTGTG 7380

GCTCCTCATC AACAATCTAT TTTGGATGCC ATCAATGCTA TCTGTAAAGA AGCGGCTGAC 7440

GAAGGTTTGG TAGATAAGTA TGTCCCTATC AAATGATCAA CTGACCTAGA AAAGGCAGCT 7500

TTTGCCAGAG CTACAGAAGC ATCTATAACC ATGGATCATA CCCGTCTTTC TAGCAAAGAT 7560

CTTTGGAGTG CCTTTCCAAC TTCTAATAGT ATAATGGGAG AAAATTTGGC ATGGAATCAT 7620

GACGGTTTTC TAAAAGCTAT TGAACAATGG CGTGCTGAAA AAGCAGATTA TGTGGAGAAA 7680

AAAATAGTGG TTCAGACAAC GGGAAATCTG GTCACTATGA GTCGCTAATT AACCCTAAAT 7740

TTACACACAT GGGGATGGCA GCTTTTAAAA ATCCTAACAA TCAATACAAA GCTATTACAA 7800

TTGCTCAAAC TCTAGGTGAT GATGCTTCTT CAGAGGAATT GGCTGGTAGA TATGGTTCTG 7860

CTGTTCAGTG TACAGAAGTG ACTGCCTCAA ACCTTTCAAC AGTTAAAACT AAAGCTACGG 7920

TTGTAGAAAA ACCACTGAAA GATTTTAGAG CGTCTACGTC TGATCAGTCT GGTTGGGTGG 7980

AATCTAATGG TAAATGGTAT TTCTATGAGT CTGGTGATGT GAAGACAGGT TGGGTGAAAA 8040

CAGATGGTAA ATGGTACTAT TTGAATGACT TAGGTGTCAT GCAGACTGGA TTTGTAAAAT 8100

TTTCTGGTAG CTGGTATTAC TTGAGCAATT CAGGTGCTAT GTTTACAGGC TGGGGAACAG 8160

ATGGTAGCAG ATGGTTCTAC TTTGACGGCT CAGGAGCTAT GAAGACAGGC TGGTACAAGG 8220

AAAATGGCAC TTGGTATTAC CTTGACGAAG CAGGTATCAT GAAGACAGGT TGGTTTAAAG 8280

TCGGACCACA CTGGTACTAT GCCTACGGTT CAGGAGCTTT GGCTGTGAGC ACAACAACAC 8340

CAGATGGTTA CCGTGTAAAT GGTAATGGTG AATGGGTAAA CTAGGCTCAG GCCATAGGTA 8400

AAGCATTCAT CTTACTTAGC AAAAAGAATG AACGATAAGA AAGAGGTTGA TGGCGAACAT 8460

TGGCCTCTTT TGATTTATAA AGATTGGATT CTTGTCGCCT CAATTTCAGA CTTTTCTATT 8520

GTAAGCTAAT ATTTTATAGC CCATTAAAAG CATAAGCGGT AATCTAATTT AAAAAATGCT 8580

GTAATTAGTC TGAAGTCCAC ACTTACTTGT TGAGATGTTA TCTCTGTTTT TTATCGTTAT 8640

AATTTACTGT ATTTTTTATA GTATGCAGAA TATTTTTAAG TATATTTCAA TAGAAATTTC 8700

TATCGATTTA TTGTATAATG ATAAGTAATT GTTGAAAAGT ACTCAGAAAA TTCCATACTA 8760

TATTATTTTT ATGTTTATAC TTTTATGCTA TAAAATATAG ATTGATATAA AGAATATAGA 8820

AAAAGCGAGG TTAATATGAG CCGAAAAAGC ATTGGTGAGA AACGCCATAG TTTCTCGATG 8880

AGAAAGTTGT CAGTGGGATT GGTATCAGTT ACTGTATCTA GTTTCTTTTT GATGAGTCAA 8940

GGGATTCAAT CGGTATCGGC CGATAATATG GAAAGTCCAA TTCATTATAA GTATATGACC 9000

GAGGGTAAAT TGACAGACGA GGAAAAATCC TTGCTGGTAG AGGCCCTTCC ACAACTGGCT 9060

GAAGAATCAG ATGATACTTA TTACTTGGTT TATAGATCTC AACAGTTTTT ACCGAATACA 9120

GGTTTTAACC CAACTGTTGG TACTTTCCTT TTTACTGCAG GATTGAGCTT GTTAGTTTTA 9180

TTGGTTTCTA AAAGGGAAAA TGGAAAGAAA CGACTTGTTC ATTTTCTGCT GTTGACTAGC 9240

ATGGGAGTTC AATTGTTGCC GGCCAGTGCT TTTGGGTTGA CCAGCCAGAT TTTATCTGCC 9300

TATAATAGTC AGCTTTCTAT CGGAGTCGGG GAACATTTAC CAGAGCCTCT GAAAATCGAA 9360

GGTTATCAAT ATATTGGTTA TATCAAAACT AAGAAACAGG ATAATACAGA GCTTTCAAGG 9420

ACAGTTGATG GGAAATACTC TGCTCAAAGA GATAGTCAAC CAAACTCTAC AAAAACATCA 9480

GATGTAGTTC ATTCAGCTGA TTTAGAATGG AACCAAGGAC AGGGGAAGGT TAGTTTACAA 9540

GGTGAAGCAT CAGGGGATGA TGGACTTTCA GAAAAATCTT CTATAGCAGC AGACAATCTA 9600

TCTTCTAATG ATTCATTCGC AAGTCAAGTT GAGCAGAATC CGGATCACAA AGGAGAATCT 9660

GTAGTTCGAC CAACAGTGCC AGAACAAGGA AATCCTGTGT CTGCTACAAC GGTGCAGAGT 9720

GCGGAAGAGG AAGTATTGGC GACGACAAAT GATCGACCAG AGTATAAACT TCCATTGGAA 9780

ACCAAAGGCA CGCAAGAACC CGGTCATGAG GGTGAAGCCG CAGTCCGTGA AGACTTACCA 9840

GTCTACACTA AGCCACTAGA AACCAAAGGT ACACAAGGAC CCGGACATGA AGGTGAAGCT 9900

GCAGTTCGCG AGGAAGAACC AGCTTACACA GAACCGTTAG CAACGAAAGG CACGCAAGAG 9960

CCAGGTCATG AGGGCAAAGC TACAGTCCGC GAAGAGACTC TAGAGTACAC GGAACCGGTA 10020

GCGACAAAAG GCACACAAGA ACCCGAACAT GAGGGCGAAg cGGCAGTAGA AGAAGAACTT 10080

CCGGCTTTAG AGGTCACTAC ACGAAATAGA ACGGAAATCC AGAATATTCC TTATACAACA 10140

GAAGAAATTC AGGATCCAAC ACTTCTGAAA AATCGTCGTA AGATTGAACG ACAAGGGCAA 10200

GCAGGGACAC GTACAATTCA ATATGAAGAC TACATCGTAA ATGGTAATGT CGTAGAAACT 10260

AAAGAAGTGT CACGAACTGA AGTAGCTCCG GTCAACGAAG TCGTTAAAGT AGGAACACTT 10320

GTGAAAGTTA AACCTACAGT AGAAATTACA AACTTAACAA AAGTTGAGAA CAAAAAATCT 10380

ATAACTGTAA GTTATAACTT AATAGACACT ACCTCAGCAT ATGTTTCTGC AAAAACGCAA 10440

GTTTTCCATG GAGACAAGCT AGTTAAAGAG GTGGATATAG AAAATCCTGC CAAAGAGCAA 10500

GTAATATCAG GTTTAGATTA CTACACACCG TATACAGTTA AAACACACCT AACTTATAAT 10560

TTGGGTGAAA ATAATGAGGA AAATACTGAA ACATCAACTC AAGATTTCCA ATTAGAGTAT 10620

AAGAAAATAG AGATTAAAGA TATTGATTCA GTAGAATTAT ACGGTAAAGA AAATGATCGT 10680

TATCGTAGAT ATTTAAGTCT AAGTGAAGCG CCGACTGATA CGGCTAAATA CTTTGTAAAA 10740

GTGAAATCAG ATCGCTTCAA AGAAATGTAC CTACCTGTAA AATCTATTAC AGAAAATACG 10800

GATGGAACGT ATAAAGTGAC GGTAGCCGTT GATCAACTTG TCGAAGAAGG TACAGACGGT 10860

TACAAAGATG ATTACACATT TACTGTAGCT AAATCTAAAG CAGAGCAACC AGGAGTTTAC 10920

ACATCCTTTA AACAGCTGGT AACAGCCATG CAAAGCAATC TGTCTGGTGT CTATACATTG 10980

GCTTCAGATA TGACCGCAGA TGAGGTGAGC TTAGGCGATA AGCAGACAAG TTATCTCACA 11040

GGTGCATTTA CAGGGAGCTT GATCGGTTCT GATGGAACAA AATCGTATGC CATTTATGAT 11100

TTGAAGAAAC CATTATTTGA TACATTAAAT GGTGCTACAG TTAGAGATTT GGATATTAAA 11160

ACTGTTTCTG CTGATAGTAA AGAAAATGTC GCAGCGCTGG CGAAGGCAGC GAATAGCGCG 11220

AATATTAATA ATGTTGCAGT AGAAGGAAAA ATCTCAGGTG CGAAATCTGT TGCGGGATTA 11280

GTAGCGAGCG CAACAAATAC AGTGATAGAA AACAGCTCGT TTACAGGGAA ACTTATCGCA 11340

AATCACCAGG ACAGTAATAA AAATGATACT GGAGGAATAG TAGGTAATAT AACAGGAAAT 11400

AGTTCGAGAG TTAATAAAGT TAGGGTAGAT GCCTTAATCT CTACTAATGC ACGCAATAAT 11460

AACCAAACAG CTGGAGGGAT AGTAGGTAGA TTAGAAAATG GTGCATTGAT ATCTAATTCG 11520

GTTGCTACTG GAGAAATACG AAATGGTCAA GGATATTCTA GAGTCGGAGG AATAGTAGGA 11580

TCTACGTGGC AAAACGGTCG AGTAAATAAT GTTGTGAGTA ACGTAGATGT TGGAGATGGT 11640

TATGTTATCA CCGGTGATCA ATACGCAGCA GCAGATGTGA AAAATGCAAG TACATCAGTT 11700

GATAATAGAA AAGCAGACAG ATTCGCTACA AAATTATCAA AAGACCAAAT AGACGCGAAA 11760

GTTGCTGATT ATGGAATCAC AGTAACTCTT GATGATACTG GGCAAGATTT AAAACGTAAT 11820

CTAAGAGAAG TTGATTATAC AAGACTAAAT AAAGCAGAAG CTGAAAGAAA AGTAGCTTAT 11880

AGCAACATAG AAAAACTGAT GCCATTCTAC AATAAAGACC TAGTAGTTCA CTATGGTAAC 11940

AAAGTAGCGA CAACAGATAA ACTTTACACT ACAGAATTGT TAGATGTTGT GCCGATGAAA 12000

GATGATGAAG TAGTAACGGA TATTAATAAT AAGAAAAATT CAATAAATAA AGTTATGTTA 12060

CATTTCAAAG ATAATACAGT AGAATACCTA GATGTAACAT TCAAAGAAAA CTTCATAAAC 12120

AGTCAAGTAA TCGAATACAA TGTTACAGGA AAAGAATATA TATTCACACC AGAAGCATTT 12180

GTTTCAGACT ATACAGCGAT AACGAATAAC GTACTAAGCG ACTTGCAAAA TGTAACACTT 12240

AACTCAGAAG CTACTAAAAA AGTACTAGGA GCAGCGAATG ATGCAGCCTT AGATAACCTA 12300

TACTTAGATA GACAATTTGA AGAAGTTAAA GCTAATATAG CAGAACACCT AAGAAAAGTA 12360

TTAGCGATGG ATAAATCAAT CAATACTACA GGAGACGGTG TAGTTGAATA CGTAAGTGAG 12420

AAAATCAAAA ATAACAAAGA AGCATTTATG CTAGGTCTTA CTTATATGAA CCGTTGGTAC 12480

GATATTAATT ATGGTAAAAT GAATACAAAA GATTTATCTA CGTACAAGTT TGACTTTAAC 12540

GGAAATAATG AGACTTCAAC GTTGGATACT ATTGTCGCAT TAGGAAATAG TGGACTAGAT 12600

AACCTGAGAG CTTCAAATAC TGTAGGTTTA TATGCGAATA AACTTGCATC GGTAAAAGGA 12660

GAAGATTCAG TCTTTGACTT CGTAGAAGCG TATAGAAAAC TGTTCTTACC AAACAAAACA 12720

AATAACGAGT GGTTTAAAGA AAATACAAAG GCATATATAG TCGAAATGAA GTCTGATATT 12780

GCAGAAGTAC GAGAAAAACA AGAATCACCA ACAGCCGATA GAAAATATTC ATTAGGAGTT 12840

TACGATAGAA TATCAGCACC AAGTTGGGGG CATAAGAGTA TGTTATTACC ACTACTAACT 12900

TTACCTGAAG AATCTGTGTA TATTTCATCG AATATGTCTA CACTTGCATT CGGTTCGTAT 12960

GAAAGATATC GTGATAGTGT GGATGGAGTT ATTCTTTCAG GAGATGCTTT ACGAACTTAT 13020

GTAAGAAATA GAGTTGATAT AGCAGCGAAA AGGCATAGAG ACCATTATGA TATTTGGTAC 13080

AATCTTCTTG ACAGTGCTTC AAAAGAAAAA CTTTTCCGTT CTGTGATAGT TTATGATGGA 13140

TTCAATGTAA AAGATGAGAC AGGAAGAACT TATTGGGCAA GGTTAACGGA TAAAAACATC 13200

GGCTCTATTA AAGAATTCTT CGGACCTGTT GGGAAATGGT ATGAGTATAA TAGTAGTGCA 13260

GGAGCGTATG CGyAtGGAAG TTTAACGCAC TTTGTGTTAG ATAGATTATT AGATGCTTAT 13320

GGAACGTCGG TTTATACTCA TGAAATGGTT CATAATTCTG ATTCTGCAAT CTACTTTGAA 13380

GGAAATGGTA GACGTGAAGG ATTGGGAGCG GAGTTATACG CACTTGGTTT ACTGCAATCT 13440

GTAGATAGTG TAAATTCTCA TATTTTAGCT TTAAATACGT TATATAAAGC AGAAAAAGAT 13500

GATTTGAATA GATTGCATAC ATATAATCCG GTGGAACGTT TCGATTCGGA TGAGGCGCTT 13560

CAAAGTTATA TGCATGGATC ATATGATGTA ATGTATACAC TTGATGCGAT GGAAGCAAAA 13620

GCGATATTAG CTCAAAATAA TGATGTTAAG AAAAAATGGT TTAGAAAAAT AGAAAATTAT 13680

TACGTTCGTG ATACTAGACA TAATAAAGAT ACACATGCAG GAAATAAAGT CCGTCCATTA 13740

ACAGATGAAG AAGTAGCTAA CTTAACATCG TTAAACTCAT TAATCGACAA CGACATCATA 13800

AATAGACGTA GCTATGATGA TAGTAGAGAA TATAAACGAA ATGGCTACTA TACTATAAGT 13860

ATGTTCTCTC CTGTATACGC AGCGCTAAGC AATTCGAAAG GTGCTCCTGG AGATATTATG 13920

TTTAGAAAAA TAGCTTATGA ATTACTTGCG GAAAAAGGTT ATCACAAAGG ATTCCTACCT 13980

TATGTTTCTA ATCAGTACGG AGCAGAAGCA TTTGCCAGCG GAAGCAAAAC ATTCTCATCA 14040

TGGCATGGAA GAGATGTTGC TTTAGTGACA GATGATTTAG TATTTAAGAA AGTATTCAAT 14100

GGTGAGTACT CATCATGGGC TGATTTCAAA AAAGCAATGT TTAAACAACG TATAGATAAA 14160

CAAGATAATC TGAAACCAAT AACAATTCAA TACGAATTAG GTAATCCTAA TAGTACAAAA 14220

GAAGTAACTA TAACAACGGC TGCACAAATG CAACAATTAA TTAATGAAGC GGCTGCGAAA 14280

GATATTACTA ATATAGATCG TGCAACGAGT CATACCCCAG CAAGTTGGGT GCATTTATTA 14340

AAACAAAAAA TCTATAATGC ATATCTTCGC ACTACAGATG ACTTTAGAAA TTCTATATAT 14400

AAATAAGATT GTAGAGTTTC ATTGTTGAGT AGTGTTTCTT GTAAGGATGA GGAGTCAGAT 14460

GACAAATCGA CTCCTTTTTC TTATGGATCG ATGTAGAGAT TTGATTGAAT GCAGATTGCA 14520

GGAATCATCT TCAACTCATC AACGACCAAT GGTGACAAGG TGGATTTCAA TCCCACAGAA 14580

AATGTTGATT TGAGAAATAA CTTTGCTAGT CTAGTAAAAT AAATACAAAA CAATCCTAGA 14640

AGATTTTTTC TGGGATTGTT TTTTGCTGAG TGGGATGCTT CAAGTTGTCT GGCTTGACTT 14700

TCTTGAGGGA AGTTATATAA TAGTTGTAAT AATTAG 14736

(2) INFORMATION FOR SEQ ID NO: 172:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11770 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

ACAGGAAAGC ACGATAGCAA TCTCTTTGGA AGATTTAAAA AATATTCCTC AAAGTTTCGC 60

TGTTGCTTAC GGTGATACGA AAGTATCTTC GATTCTCTCT GTCTTGCGTG CTAATTTAGT 120

AAATCATTTG ATTACAGACA AAAATACAAT TTTAAAAGTT TTGGAAGAAG ATGGGGATTT 180

GACTTTTAGA GAGATTCTAG GTGAGTGAAA ATGATAGACT GATTCAGTTT ATCGTTTTTC 240

TTTTTAGTTG ATTGCACATT TGTGCTTATA TAAACAAAAA TAGTTTATCT GTTGTTTTTG 300

GATTGACAAC TTTATTATGT AGTTGTATTC TATAGTTACA AAAGAAAATT TTAAAATTTC 360

AAATGAAAAA AGCTTTTTAC ATAGTGAAAT GAGGAGGAAT TTATGGAAAT GATTGTTCCA 420

GATCAAATTA TCATGGGTTT AATTTTATAT GCTGGTGATG CGAAACAACA TATTTATAAA 480

GCGTTAGATT ACATAAAAAA TGGTACATGT GAACGGTGTG AAGAAGAAAT ACAGTTAGCT 540

GATGCAGCCT TATTAGAAGC TCATAATCTA CAAACAAAAT TTTTGGCACA GGAAGCGTCT 600

GGTACAAAGA CAGAAATTAC AGCTCTCTTT GTTCATTCAC AAGATCATCT CATGACCAGT 660

ATGACGGAGA TTAATTTAAT CAAAGAAATT ATTAGTTTGA GAAAAGAACT TCATAAAAAA 720

TAATACTAGA GTATTATCAT TGTTATTAAC ATAGAGGAGG AAAACATAAT GGTGAAGATT 780

GGTTTGTTTT GTGCAGCAGG TTTTTCTACT GGTATGCTTG TAAATAATAT GAAAATTGCA 840

GCGCAATCTA GTGGAGTTGA GGCAGAAATA GAGGCGTTTT CTCAGTCTAA ATTAGCGGAT 900

TATGCGCCAA ATATAGATGT TGCACTATTG GGTCCACAAG TTGCTTATAC ATTAGATAAA 960

TCAAAAGAAA TTTGTGATAA GTGTGATGTT CCGATAGCTG TTATTCCGAT GATGGACTAT 1020

GGTATGTTAG ATGGGAAAAA AGTATTAGAT TTGGCCCTAT CTTTGATTAG TGGGTAAGAA 1080

AAGGAGATTT ATTATGTCAA AGATGGATGT TCAGAAAATC ATTGCACCGA TGATGAAGTT 1140

TGTGAATATG CGTGGCATTA TAGCTCTAAA AGATGGGATG TTAGCAATTT TGCCATTGAC 1200

AGTAGTTGGT AGTTTGTTCT TGATTATGGG ACAATTGCCG TTCGAAGGAT TAAATAAGAG 1260

CATTGCTAGT GTTTTTGGAG CTAATTGGAC AGAGCCGTTT ATGCAAGTAT ATTCAGGAAC 1320

TTTTGCTATT ATGGGTCTAA TTTCTTGTTT TTCAATTGCC TATTCTTATG CTAAGAATAG 1380

CGGAGTAGAG GCTTTACCAG CTGGAGTTCT ATCTGTATCT GCATTCTTTA TTTTGCTAAG 1440

ATCATCTTAT ATCCCTAAAC AAGGTGAGGC GATTGGGGAC GCTATTAGTA AAGTTTGGTT 1500

TGGAGGCCAA GGAATTATCG GTGCTATCAT TATAGGTTTG GTAGTAGGAA GTATTTATAC 1560

CTTCTTTATA AAGAGAAAAA TTGTTATTAA GATGCCAGAA CAAGTTCCAC AAGCTATTGC 1620

CAAACAGTTT GAAGCAATGA TTCCAGCATT TGTAATTTTC TTATCTTCTA TGATTGTATA 1680

TATTTTAGCG AAGTCATTGA CTAATGGCGG AACATTCATA GAAATGATTT ATTCTGCTAT 1740

TCAAGTTCCG TTGCAAGGTT TAACTGGATC TTTGTATGGT GCTATTGGAA TTGCATTCTT 1800

TATATCATTT TTGTGGTGGT TTGGTGTTCA TGGGCAATCG GTAGTAAATG GAGTAGTGAC 1860

AGCTCTGCTT TTATCTAATC TTGATGCTAA TAAAGCTATG TTAGCCTCTG CTAATCTATC 1920

ATTAGAAAAT GGTGCACATA TTGTTACTCA ACAATTTTTA GATTCATTTT TAATTCTATC 1980

AGGTTCAGGG ATTACGTTTG GTCTTGTAGT TGCCATGCTT TTTGCAGCAA AATCAAAACA 2040

ATACCAAGCC TTAGGAAAAG TTGCAGCTTT TCCAGCAATA TTTAACGTAA ATGAGCCAGT 2100

TGTATTTGGA TTTCCGATTG TCATGAATCC AGTTATGTTT GTACCTTTCA TTCTTGTTCC 2160

TGTACTTGCA GCTGTGATAG TATATGGAGC TATTGCAACA GGTTTCATGC AGCCATTCTC 2220

AGGGGTAACA TTGCCTTGGA GTACACCAGC TATTTTATCA GGATTTTTGG TGGGTGGATG 2280

GCAAGGAGTT ATTACTCAGC TGGTGATATT AGCGATGTCT ACATTGGTTT ATTTTCCATT 2340

CTTTAAAGTA CAGGATCGTT TAGCTTACCA AAATGAAATC AAACAATCTT AGAGGTATTT 2400

GTGTGTTACT GTTAAACTCA CACATTTGTG CTAAAAATTA GAGAGTTAAA ATTTTTCTAG 2460

TTAAAAGCTT GAAAATTTCT ATAAAAATCG GTATTATATT TTCGAAAGAA ATAAAAATAT 2520

TTTCGAAAGA AAGGTGCTTA CGATGGTAAA TACAGAAGTA GCAAGAACAA CAATCAAGAC 2580

AGAATATTTT GGCAGCCTTA CTGAAAGGAT GAACAAATAT CGAGAAGATG TTTTAAATAA 2640

AAAACCTTAT ATTGATGCTG AGAGAGCAGT TCTAGCAACA CGCGCCTATG AACGATACAA 2700

GGAACAACCT AATGTCCTAA AACGTGCATA TATGCTGAAA GAAATTTTGG AAAATATGAC 2760

TATCTATATT GAAGAAGAAT CTATGATTGC GGGAAATCAA GCTTCTTCCA ATAAAGATGC 2820

TCCTATTTTT CCGGAATATA CGCTAGAATT TGTTCTCAAT GAGTTGGATC TTTTTGAAAA 2880

GCGTGATGGA GATGTTTTCT ATATTACAGA AGAAACAAAA GAACAACTTA GAAGTATTGC 2940

TCCGTTTTGG GAAAATAATA ATTTACGTGC TAGAGCTGGT GCCTTATTAC CTGAAGAAGT 3000

GTCTGTTTAT ATGGAAACAG GATTCTTCGG TATGGAAGGT AAGATGAATT CTGGAGATGC 3060

TCACTTAGCA GTTAACTATC AGAAACTTTT GCAATTTGGT TTAAGAGGTT TTGAAGAGCG 3120

GGCTCGTAAA GCAAAAGTAG CTCTAGATTT AACAGATCCA GCAAGTATTG ATAAATATCA 3180

TTTTTACGAC TCTATATTTA TCGTAATCGA TGCTATTAAA GTATATGCAA AGCGCTTTGT 3240

TGCTCTTGCT AAAAGTTTAG CCGAAAATGC AAATCCTAAA CGTAAGAAAG AATTACTTGA 3300

GATTGCAGAT ATTTGCTCTA GAGTCCCATA TGAACCGGCA ACTACTTTTG CAGAAGCTAT 3360

TCAATCAGTT TGGTTTATTC AATGTATTTT ACAAATTGAA TCTAATGGCC ACTCTCTTTC 3420

ATATGGCCGT TTTGATCAAT ATATGTATCC ATATATGAAG GCTGATTTAG AAAGTGGTAA 3480

AGAAACAGAA GATAGCATTG TTGAACGTCT GACAAATCTT TGGATTAAGA CAATTACAAT 3540

TAATAAGGTT CGCAGTCAAT CACATACATT TTCTTCAGCA GGAAGTCCTT TATATCAAAA 3600

TGTTACAATT GGTGGACAGA CTCGAGATAA GAAGGATGCT GTTAACCCAT TATCTTATTT 3660

GGTATTAAAA TCAGTTGCAC AAACCCATCT ACCGCAACCT AATCTAACTG TACGTTACCA 3720

TGCAGGTTTA GATGCTCGTT TCATGAATGA GTGTATTGAA GTGATGAAAC TTGGTTTTGG 3780

TATGCCTGCA TTTAATAATG ATGAGATTAT TATTCCTTCT TTTATTGCAA AAGGAGTATT 3840

GGAAGATGAT GCTTATGATT ACAGTGCCAT TGGATGTGTT GAAACGGCAG TTCCAGGGAA 3900

ATGGGGCTAT CGTTGCACAG GTATGAGTTA TATGAACTTC CCTAAGGTTC TACTTATCAC 3960

GATGAATGAT GGAATTGATC CGGCTTCGGG TAAACGGTTT GCACCAAGCT TTGGTCGTTT 4020

TAAGGATATG AAGAACTTTT CTGAATTAGA AAATGCTTGG GATAAAACAC TAAGATATTT 4080

GACACGAATG AGTGTTATTG TTGAAAATTC TATTGATTTA TCATTGGAAC GAGAAGTTCC 4140

TGATATTCTA TGTTCAGCAT TGACTGATGA TTGTATTGGT CGTGGAAAAC ACCTTAAAGA 4200

AGGTGGAGCA GTATATGATT ATATATCAGG ATTGCAAGTT GGAATTGCAA ATTTGTCGGA 4260

TTCATTAGCT GCAATTAAAA AATTGGTGTT TGAGGAAGAA CGTATAAGCC CAAGTCAGCT 4320

TTGGCATGCA CTGGAAACAG ATTATGCCGG AGAAGAAGGT AAGGTCATTC AAGAAATGTT 4380

GATTCATGAT GCACCTAAGT ATGGTAATGA TGATGATTAT GCTGACAAAT TGGTTACTGC 4440

TGCTTATGAC ATTTATGTTG ATGAAATTGC TAAATATCCT AATACACGTT ATGGAAGAGG 4500

GCCTATTGGA GGAATTCGTT ATTCAGGAAC ATCTTCTATC TCAGCCAACG TAGGGCAGGG 4560

ACGTGGAACA TTAGCAACTC CAGATGGACG CAACGCGGGT ACACCGTTAG CAGAGGGTTG 4620

TTCACCATCA CATAATATGG ATCAACACGG CCCTACATCT GTTTTAAAAT CTGTTTCAAA 4680

ATTACCAACA GATGAAATCG TAGGTGGGGT TCTCTTAAAT CAGAAAGTAA ATCCTCAAAC 4740

GTTAGCCAAA GAAGAAGATA AATTAAAACT AATTGCTTTG TTACGAACAT TCTTTAATCG 4800

TTTACATGGG TACCATATTC AATACAATGT TGTTTCCAGA GAGACGCTGA TTGACGCTCA 4860

GAAACATCCT GAAAAACACA GAGACTTAAT TGTTCGTGTT GCAGGATACT CTGCATTCTT 4920

CAATGTTCTT TCTAAGGCAA CCCAAGATGA CATTATAGGA CGTACTGAGC ATACTTTGTA 4980

AAATAAAGAG GTTCTTTTTA TGGAATTTAT GCTTGACACA TTAAATTTAG ATGAGATTAA 5040

AAAGTGGTCT GAAATTTTGC CGCTAGCTGG GGTAACTTCA AATCCCACTA TTGCAAAAAG 5100

AGAGGGTTCT ATTAATTTTT TTGAACGAAT CAAAGATGTA AGAGAATTGA TTGGCTCTAC 5160

ACCCTCTATT CATGTTCAGG TGATTTCTCA AGATTTTGAA GGCATCTTAA AGGATGCTCA 5220

TAAAATTCGA AGACAAGCAG GAGATGATAT ATTTATCAAA GTACCTGTTA CTCCAGCTGG 5280

ATTACGTGCA ATAAAGGCGC TAAAAAAAGA GGGCTACCAT ATCACTGCAA CAGCTATTTA 5340

TACAGTTATT CAGGGATTAT TAGCTATCGA AGCAGGAGCG GATTACCTAG CTCCATATTA 5400

TAATAGAATG GAAAATCTGA ACATTGATTC AAATTCTGTC ATTCGTCAAT TAGCTCTTGC 5460

TATTGATAGA CAGAACTCTC CTAGTAAGAT TTTAGCTGCA TCCTTTAAAA ATGTAGCACA 5520

AGTAAATAAT GCTTTAGCTG CAGGTGCGCA TGCTGTTACA GCAGGAGCGG ATGTTTTTGA 5580

ATCAGCTTTC GCCATGCCAT CTATCCAAAA GGCGGTTGAT GATTTTTCTG ACGATTGGTT 5640

TGTTATTCAA AATAGTCGTT CCATTTAGAT AGAGAGGAAA TACATATGAG AATTTTTGCT 5700

AGTCCTTCTA GATATATTCA GGGGGAAAAT GCCTTGTTTG AAAATGCCAA ATCAATTTTG 5760

GATTTGGGAA ATTGCCCTAT TCTATTATGC GATCAGTTGG TTTATGATAT TGTTGGAAAA 5820

CGATTTGAAG ATTACCTACA TAGGTATGGT TTCCATATTG TTCTGGCGCT ATTTAATGGT 5880

GAAGCTTCTG ACAATGAAAT CAATCGAGTT GTTGCCTTGG CTGAGAAAGA AAATTGTGAT 5940

AGTATTATCG GTCTTGGTGG GGGAAAGACG ATTGATAGCG CAAAAGCTAT TGCAGATTTG 6000

ATTGAAAAGC CTGTTATTAT TGCTCCAACA ATTGCATCGA CCGACGCACC TGTATCTGCT 6060

TTATCTGTTA TTTATACAGA TGAAGGTGCA TTTGATCATT ATCTATTTTA TTCTAAAAAT 6120

CCAGATTTAG TTTTGGTTGA TACAAAAGTT ATTTCACAAG CCCCTAAGCG TTTATTAGCG 6180

TCTGGTATTG CAGATGGTTT AGCAACTTGG GTTGAGGCGC GTGCGGTTAT GCAGGCAAAT 6240

GGAAAAACTA TGTTGGGACA ACAGCAAACA TTGGCTGGAG TTGCAATTGC GAAGAAATGT 6300

GAAGAAACGC TGTTTGCAGA TGGTTTACAG GCTATGGCAG CTTGTGAAGC TAAAGTGGTG 6360

ACACCAGCAT TAGAAAATAT TGTTGAAGCT AATACTTTAT TGAGTGGTCT AGGTTTTGAA 6420

AGTGGAGGAT TAGCTGCGGC GCATGCAATT CATAATGGTT TTACTGCATT GACAGGTGAC 6480

ATTCATCATT TAACACATGG TGAAAAAGTA GCTTATGGAA CTTTAGTACA ACTATTATTG 6540

GAAAATAGAC CTAAAGAAGA ACTTGATAAG TATATTGAGT TTTACAAAAA AATTGGTATG 6600

CCAACAACTC TAAAAGAAAT GCATTTGGAT CAAGTTGGAT ATGATGATTT AATAAAAGTT 6660

GGTAAACAAG CAACTATGGA GGGTGAGACA ATTCATCAGA TGCCGTTTAA GATTTCGCCT 6720

TCAGATGTTG CTCAAGCTAT TATCGCTGTA GATGCCTATG TAAATTCAAA ATAAACAATA 6780

AGGACTACTG TTTTCCAAAT GGTAGTCTTT TATTGATCCC TGTATTGAAT TCTATAGAAG 6840

ATTGAAATAG GATGAGAACA AATCGATTGG GAAAGTAAAA TTAATTTCTA TAAATGTTTT 6900

AGCAATTGTT TCGTACTATT TCAGATTCAG TCTACTATAT GTTCTTCATA AATCAAAAAG 6960

CGACATAGGT TGTCGGCTAT TTATTGTGAA TACATTAATT AGCATTCCAG TTTTATCTTC 7020

GGTCTAAAAT AAGTATTTTG TGCTATACGA GATAAGCTTC TTGACTTACT CCTTGATTTA 7080

CTGCATAACA ATGGGATAAA AAGTGGGAGA TAGAGCAATT CATAGTCATC AAAATTAATG 7140

AGATACAGTA TACAGTTTTT CCTTTAAACA CATTTCAAAT TCCCTCAAAA ATGGTATAAT 7200

AGTAACATCA CAAAATTGGA GAGAGACCAT GAGTTTTTAC AATCATAAAG AAATTGAGCC 7260

TAAGTGGCAG GGCTACTGGG CAGAACATCA TACATTTAAG ACAGGAACAG ATACATCAAA 7320

ACCTAAGTTT TATGCGCTTG ATATGTTCCC TTATCCGTCT GGAGCTGGTC TGCACGTAGG 7380

ACACCCAGAA GGTTATACTG CAACCGATAT CCTCAGTCGT TACAAACGTG CGCAAGGCTA 7440

CAATGTCCTT CACCCAATGG GTTGGGATGC TTTTGGTTTG CCTGCAGAGC AATACGCTAT 7500

GGATACTGGT AATGACCCAG CAGAATTTAC AGCGGAAAAC ATTGCCAACT TCAAACGTCA 7560

AATTAATGCG CTTGGATTTT CTTATGACTG GGATCGTGAA GTCAACACAA CAGATCCAAA 7620

CTACTACAAG TGGACTCAAT GGATTTTCAC CAAGCTTTAC GAAAAAGGCT TGGCCTATCA 7680

AGCTGAAGTG CCAGTAAACT GGGTTGAGGA ATTGGGAACT GCCATTGCCA ATGAAGAAGT 7740

GCTTCCTGAC GGAACTTCTG AGCGTGGAGG CTATCCAGTT GTCCGCAAAC CAATGCGCCA 7800

ATGGATGCTC AAAATCACGG CTTACGCAGA GCGCTTGCTC AATGACTTAG ATGAACTAGA 7860

TTGGTCAGAG TCTATCAAGG ATATGCAACG CAACTGGATT GGTAAATCAA CTGGTGCCAA 7920

TGTAACTTTC AAAGTAAAAG GAACAGACAA GGAATTTACA GTCTTTACTA CTCGTCCGGA 7980

CACACTTTTC GGTGCGACTT TCACTGTCTT GGCTCCTGAA CATGAATTAG TAGACGCTAT 8040

CACAAGTTCA GAGCAAGCAG AAGCTGTAGC AGACTATAAA CACCAAGCCA GCCTTAAGTC 8100

TGACTTGGCT CGTACAGACC TTGCTAAAGA AAAAACAGGG GTTTGGACTG GTGCTTATGC 8160

CATCAACCCT GTCAATGGTA AGGAAATGCC AATCTGGATT GCAGACTATG TCCTTGCTAG 8220

TTATGGAACA GGTGCGGTTA TGGCTGTGCC TGCCCACGAC CAACGTGACT GGGAATTTGC 8280

CAAACAATTT GACCTTCCAA TCGTCGAAGT ACTTGAAGGT GGAAATGTCG AAGAAGCTGC 8340

CTACACAGAG GATGGCCTGC ATGTCAATTC AGACTTCCTA GATGGATTGA ACAAAGAAGA 8400

CGCTATTGCC AAGATTGTGG CTTGGTTGGA AGAAAAAGGC TGTGGTCAGG AGAAGGTTAC 8460

CTACCGTCTC CGCGACTGGC TCTTTAGCCG TCAACGTTAC TGGGGTGAGC CAATTCCAAT 8520

CATTCATTGG GAAGATGGAA CTTCAACAGC TGTTCCTGAA ACTGAATTGC CGCTTGTCTT 8580

GCCTGTAACC AAGGATATCC GTCCTTCAGG TACTGGTGAA AGTCCACTAG CTAACTTGAC 8640

AGATTGGCTT GAAGTGACTC GTGAAGATGG TGTCAAAGGT CGTCGTGAAA CCAACACTAT 8700

GCCACAATGG GCTGGTTCAA GCTGGTACTA CCTCCGCTAT ATTGACCCGC ACAATACTGA 8760

GAAATTGGCT GATGAGGACC TCCTCAAACA ATGGTTGCCA GTAGATATCT ACGTGGGTGG 8820

TGCGGAACAT GCTGTACTTC ACTTGCTTTA TGCTCGTTTC TGGCATAAAT TCCTCTATGA 8880

CCTCGGTGTT GTTCCGACTA AGGAACCATT CCAAAAACTC TTTAACCAAG GGATGATTTT 8940

GGGAACAAGC TACCGTGACC ACCGTGGTGC TCTTGTGGCA ACCGACAAGG TTGAAAAACG 9000

TGATGGTTCC TTCTTCCATG TAGAAACAGG GGAAGAGTTG GAGCAAGCGC CAGCCAAGAT 9060

GTCTAAATCG CTCAAGAACG TTGTTAACCC AGACGATGTG GTGGAACAAT ACGGTGCCGA 9120

TACCCTTCGT GTTTATGAAA TGTTTATGGG ACCACTCGAT GCTTCGATTG CTTGGTCAGA 9180

AGAAGGTTTG GAAGGAAGCC GTAAGTTCCT TGACCGAGTT TACCGTTTGA TTACAAGTAA 9240

AGAAATCCTT GCGGAAAACA ATGGTGCTCT TGACAAGGTT TACAACGAAA CAGTCAAAGC 9300

TGTTACTGAG CAAATTGAGT CTCTCAAATT CAACACAGCT ATTGCCCAAC TTATGGTCTT 9360

TGTCAATGCT GCTAACAAGG AAGATAAGCT TTATGTTGAC TATGCCAAAG GCTTTATTCA 9420

ATTGATTGCA CCATTTGCAC CTCACTTGGC AGAAGAACTC TGGCAAACAG TCGCAGAAAC 9480

AGGTGAGTCA ATCTCTTATG TAGCTTGGCC AACTTGGGAC GAAAGCAAAT TGGTTGAAGA 9540

TGAAATTGAA ATTGTCGTCC AAATCAAAGG AAAAGTTCGT GCCAAACTCA TGGTTGCTAA 9600

AGATCTATCA CGTGAAGAAT TACAAGAAAT CGCTTTAGCT GATGAAAAAG TCAAAGCAGA 9660

AATTGACGGT AAGGAAATCG TGAAAGTAAT TGCGGTACCG AATAAACTCG TTAATATCGT 9720

CGTTAAATAA CGAGTTTATT AGCTCTATCT GCCACCTTCA ATAGTCCACT GGACTATTGA 9780

AsCCAACTAA ATTAGTTAAC ATTGTTGTGA AATAAGATAG GAGTCCTTCA GAGTAGAATC 9840

1100

TGGAGGATTT TTTGAATCTT CTTATGAAAG TATGATATAC TATGGGCAAC TATAAAGTTT 9900

GAAAAGTGAA ATAAGGAGAA TAAGATGCCA GTAAATGAAT ATGGTCAAAT GATTGGGGAG 9960

TCAATGGAAG CTTATACTCC AGGTGAATTG CCTTCTTTTG ATTTCTTAGA AGGGCGTTAT 10020

GCTAGGATAG AGGCTCTTTC AGTGGAAAAG CATGCGGAGG ATTTATTAGC TGTTTATGGC 10080

CCTGATACGC CTCGGGAGAT GTGGACCTAC CTCTTTCAGG AGTCAGTAGC AGACATGGAG 10140

GAACTGGTCA GCCTTTTAAA TCAGATGTTG GCTCGTAAGG ACCGTTTTTA TTATGCAATC 10200

ATAGACAAGG CAACTGGTAA GGCTTTGGGA ACTTTTTCCC TCATGCGAAT TGATCAGAAT 10260

AACCGAGTAA TAGAAGTGGG AGCTGTCACT TTTTCTCCAG AGCTCAGGGG GACACGGATA 10320

GGAACAGAAG CCCAGTATCT CTTGGCTTGC TATGTCTTTG AGGAGCTTAA CTATCGTCGC 10380

TATGAGTGGA AATGCGATGC TCTTAACCTG CCATCCAGAC GAGCAGCGGA ACGTTTGGGA 10440

TTTATTTATG AAGGAACCTT CCGTCAGGCA GTGGTTTATA AGGGGCGTAC AAGAGATACG 10500

GATTGGTTGT CTATGATTGA TAAGGACTGG CCTCAAGTCA AAGCTCGATT GGAAATATGG 10560

TTGCGTCCTG AAAACTTTGA TAAAAATGGA CGACAGCACA AGAGCTTGAG AGAACTTTAA 10620

GAGGTGTTGA GATGATTACT ATTAAAAAGC AAGAAATTGT CAAGCTAGAG GATGTTTTGC 10680

ATCTCTATCA GGCTGTCGGT TGGACAAACT ATACCCATCA AACAGAGATG CTGGAGCAGG 10740

CCTTATCTCA TTCATTAGTA ATTTATCTGG CACTTGATGG TGATGCTGTG GTGGGCTTGA 10800

TTCGTTTGGT TGGAGATGGT TTTTCATCAG TTTTTGTACA GGATTTGATT GTTTTGCCTA 10860

GCTATCAGCG TCAAGGGATT GGTAGCTCCT TGATGAAAGA GGCTTTAGGA AATTTTAAAG 10920

AGGCCTATCA AGTCCAGCTG GCGACAGAAG AGACAGAAAA AAACGTGGGA TTTTATCGTT 10980

CTATGGGCTT TGAAATCTTA TCCACCTATG ACTGTACAGG AATGATTTGG ATAAACAGAG 11040

AAAAATAAAA AAACTTGTTT GTTCTTAAGC AAAGTTTAAG GATGGTCTAG TATCATATAG 11100

TCATTAAATA AAGACCTCCT AACTTTATTT AATAAAATCC TAAACTTTTT TCATCACAAT 11160

CTCCTAATGA AGCCACCCAA TCAGGTGGCT TTTTTGCGGT ACGACGGGCA TGTCGTATAT 11220

CTGAGGTGTA AGTCCTCAGC CTGACTATCG TGAGGTAGCA GGGAGAGGAA GGGATAGCGA 11280

AATCGTGGCT CTACGAACAG GAACGTGATA GTAAGGCGTA TATAGCGGAT AAGGAGGCTT 11340

CAAACTCTAA AGTCCAAAAA GGTAGTCGTA ACCTATATGT GTAAATCACG AGAGTAATTG 11400

AATTCGGACT AAGGTTTGTG TGAAAAAGAT AAATCTTTCT AGAGTCTAAA GACTCTGCGT 11460

CAGATTTCCT ATTTTCACTG TAACCTTTTA ACGTCCTCAT ATCTTGTATA AACGAGGAAA 11520

GATGTACGAC TTATCCCGTG AGGTTTCATG AGCGCTGAAA GCGTAGTAAC AACGAATCAT 11580

GAGAAGTCAG CCGAGCCCAT AGTAGTGAGG AAACTTCCGT AATGGAAGTG GAGCGAAGGG 11640

GTGAATACTC AAACAGTCTG GGGAGAGACT GTTTGAGGTC TGTCGCTAGA AAGAGAAAAC 11700

GACAGATCGA AGTAATCCTA CTTCACTTGT GTCTGTAAAA TGAGTGGTCT GATAGAACTG 11760

GACTTTGAGG 11770

(2) INFORMATION FOR SEQ ID NO: 173:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4185 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

CGCGAAACTA CTTTCTTAGT ATAACACTTT CAGAATCATT GTCAATAGAA ATGACTTGAT 60

TTTTTCAATT TTTTCAAGCT ATTTCCAAGG GTTGTAAAAT CGTCCCTGAT TCTGCAAGAT 120

AAGTAGTAAA CTAACTACTA AAAACAAGGT TGCCAAGAGC AAGGTAATAT AGTCTCCTTT 180

TTTCAAGGCC TGATAACTAT ACCATGTGCG TTTTTTCTCT TTCCCAAAGC GGCGAACTCC 240

ATGGCAGTCG CAATGGTATC AATGCGTTCT AGCGAGCTAA AAATCAAGGG CGTAATAATG 300

AGCAGATTGC CTTTGATTCG TTGCATAAGA GAAGCTTTCT TGGATAATTC CATCCCACGC 360

GCCTCCTGAG ACATCTTGAT AGTAAAGAAT TCTTCCTGCA AATCTGGAAT ATAGCGCAAG 420

GTCAGGCTGA CAGAATAAGC AATCTTATAG GGCACACCAA TTTGATTTAA ACTGGAAGCA 480

AACTGACTAG GATGGGTTGT CATCAAAAAG ATAATAGCCA GAGGAATGGT GCAAAGATAC 540

TTAATGGCCA AATTTAGCAG ATAAAAGAGC TCCTGGCTGG TTAGAGTGTA GACACCGATT 600

CCCTGCCAAA TCACACTTCT CTCTCCATAA AGTCCAACCC CATACTCGGG AGAAAAGAGA 660

TAGACCATCA AAACGTTTAA AACGGCAAAT ATCGTCGCAA AAACGGCTAC AAAGGAAACA 720

TCTTTAAAGC GAATTTCTGA TAAATAGAGG AGAAAGACTG AAAAGATGGC AATCAGCAAG 780

AGCATTCTGG TATCATAGCT AATCATGGCC GCCAATGATA CCAGAATGAA AAAGAGAAGT 840

TTCCCAGCTC CTGACAAGCG ATGAATCACA GTATCTCTAT GCTGGTAACC GATTAATTTA 900

GCTTGCATCC CTCTCTCCTT TCTTTGTAAA ATGCCGTTAA ATCCAGTGGA TCCACATCTA 960

GTTTCTTAGC CAAGTTAAAG ATGGAGGTTT CTTTTAGATT GGCTTTTACT AACAGCTCAG 1020

GATCGCTCAA CAGACTGGCT GGAACAGTAT CGGCAATCAA TTCTCCATCC ACCATGACAA 1080

GGACCCGGTC TGAATAATCC AGCATCAATT GCATATCATG GGTAATCATG ACAATGGTAT 1140

GCCCTTTTTG ATGTAACTCT TCGAGAAATT CCATAATCTC AGTATAGTTC TTCTGATCTT 1200

GACCTGCAGT CGGTTCATCT AGGAGAATAA TTTCAGCTCC TAAGACCAAA ATTGAAGCAA 1260

TGGTGACACG TTTTTTCTGA CCAAATGACA GGGCAGAAAT AGGCCAATTA CGGAATTCAT 1320

AAAGTCCACA GATTTTCAAG GTTTCATATA CTCTCGTTTC AATTTCCTTC TCATCCACAC 1380

CTCGCAAACG GAGCCCTAGA GCCACCTCAT CAAAAATCAT ATTGGTTGAA ATCATTTGAT 1440

TAGGATTTTG TAGCACATAT CCTACTCGTT CCGCCCGCTC TGCAACAGAA TCGCCTTTTA 1500

TATCCTGTTT TTCCCAAAGA TAGCGTCCTT CCGTCTGAAT AAAGCTACTT ATAGCCTTGG 1560

CTAGAGTTGA TTTCCCTGCT CCATTTTTTC CGACAATAGC AATCTTTTCA CCCTTTTTAA 1620

TATCTAAATG TAGGGATTTT AAAATCGGTC TATCATCATA AGAAAAAGAT ACTTCCTCTA 1680

GTCTAAAGAG TGACTGCAAT GCTGGGGTTT CTTTTGCCAG TTCATTCTGC AACTGAACCT 1740

GACCTTTTGA GATAGACAAG TTATCCAGAT TCGCTAATTG TTCTTCCTTG ACTAAGTCCA 1800

CACCTAATTG ACGGAGAGTC GTTAGATAAA GGGGTTCTCG AATTCCATTT TGAGTCAATA 1860

AATCAGTCGC AAGCAACTGG TCAGGGCTCC CATTAAAAAG GATACGACCA TCGTTTATCA 1920

AGACAATCCG ATCCACAGGG CGATGCAGAA CGTCCTCCAA ACGGTGCTCG ATAATAAGAG 1980

TCGTCGTCCC CTCTTCCTTA TGAATCTGGT CAATCAATTC GATAATATCC TGACCTGACT 2040

TGGGATCTAG ATTGGCGAGT GGCTCATCAA ACAAGAGAAT CGGACTTTCA TCAATCAAGA 2100

CACCAGCCAG ACTGACTCGC TGCTTTTGTC CACCTGACAA ATCCTGAGGA CGCTGATCCA 2160

GTAAAGGAAG AAGGTCCAGC TTTTCAGCCC ATTTATAAAC ACGACCTTTC ATCTCATCTA 2220

GGGCTGTCAC ATCATTTTCC AGAGCAAACG CCAAATCTTC TGCCACAGAC AAGCCAATAA 2280

ACTGCCCATC TGTATCCTGC AAAACTGTGC TAACCAGATG AGACTTATCA TAGATGCTCA 2340

TATCAAAGGC TACTTGACCC TTTATCAAAA ATTCTCCATA TGTCTGACCC TTGTAAATAT 2400

TGGGAATAAT CCCATTCAAA CACTGACCCA AGGTAGATTT ACCTGACCCA GATGGTCCAA 2460

CAATTAAGAC TTTCTCTCCC TTGTAAATGG TCAAGTCTAT CCCTTGCAAG GTCGGTTCTT 2520

GTTGTGTTTC ATACCGGAAA GAGAAATCCT TCCACTCAAT TaTAGCTTCT TTCATCTTAC 2580

TCTCTTCATT CGCTTCTTAG ACTTCTATTT TATCATAAAT CAAGCCCTTC TTGCAGTCTC 2640

TCCTCTTAAA ATCTTAGCGC CAAAAAGATT CCTATCCTAG CTTACTTGCC TAACTAATCT 2700

ATAAACATCG AAAAAGACTA GTTGCCCAGC CTTCCCCATC ATTTTATACT CTTCGAAAAT 2760

CTCTTCAAAC CACGTCAGcT TCGCCTTGCC GTAGGTATGG TTACTGACTt CGTCAGTTTC 2820

ATCTACAACC TCAAAACCAT GTTTTGAGCc TGCTTCGTCA GTTCTATCCA CAATCTCAAA 2880

ACACTGTTTT GAGCAACtGC GGCTAGCTTC CTAGTTTGCT CTTTGATTTT CATTGAGTAT 2940

TAGTCCTTTT TCAAACTTCC TGCACGAGTT TGGGTTCCTG CATAGGCAAG TAAGAGAAGA 3000

GTTCCTGCAA TAGCTACAGA TACACCATTG GCAATTCCCG CAACAATCCC TTGTGCAAAT 3060

ACTTTTTCTG CCGCTTCTTG ATAAATCACA ACATCTCCAA GTGGTGCCAA GACACCCCAA 3120

ACAAGGGCAT TTGCAAGTAG TTGAATGAGA TTAAAAATAA GAATATCTTT CCAGTCAAAA 3180

ACACCATTGA TCACGCGAAC GTACTTTCTA AAAAGTCCCA CAACTAAACC AAAGAGTCCG 3240

CTAGCGATAA TCCAAGTCCA CCATAGACCA TAACCAACAA GAGAGTCCTT GATTGCATGA 3300

CCAATCAACC CGACAAGCAA ACCGATAATC GGTCCAAAAA TAATAGAAAG TAGCGCTTGT 3360

ACCGCATACT GAAGCTGGAT GCTTGTATTT GGAACAGGGG TTGGAATGTT GATCATCCCG 3420

ATGACGACAA AGAGGGCAGC GCCAATTCCG ACAGCAACAA CTTGTTTAAT TGTAAATTTG 3480

ATTTCCATAC TATTCTCCTA TTTTATCCTT CTATTTTCTT TATTTCAATG GTCCAAGATG 3540

AACCGACACC TACATTATAG GCCTTGGCAA AGGAACCTTG GTTGATAGCC AAACCTAAAC 3600

GATAGAGAGA GTTGATGTAA AGGATGGGTT GCCCAATTCT CACATCTGCA AATGATTTGC 3660

CATAGACAAC CTGATTTTGA TAGACCAGCA TATCAGCATG ATAGATGGTC ACTTCAAAAC 3720

GATCACCAAA TTCTGGTTCC AGCTTGTAAA ATTCTTCCCG TGTGATAGAG GTCCAAAGCG 3780

AACCGAAACG CACATCCAGA ATATCAATGG CTCCCTTCAC CAGATGATCT TCTATGATGG 3840

TCGCTACGAC TGGAAGCTCT ACAATCTGTT CCACACTGAG CTCTGGCCCT ACTTCCTCAA 3900

AAGTAATGTG ACCACTGGCC AGTTTAGCAC CAGTATAGGC ATAGACATCA CGACCGTGGA 3960

AGGTATAAGA ATGCTCTGTG TTTTGACGCC TATTGGCCAC CTCAGAAATC TCACGAATGG 4020

CTACAATGCC AACGTGTTTC TTGATAAAGG AAAGCGTCCC ATTATCTGGC GTGACAATGT 4080

ATTGATTTTT TGCAGTCTTG GCAACTACAC TCTTACGTTT CGAACCGACA CCTGGATCGA 4140

CAACCGATAC AAACGTCGTT CCCTCAGGCC AGTAATCCAC CGTCT 4185

(2) INFORMATION FOR SEQ ID NO: 174:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2069 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

TGATAGAGTT AAAGCCGCTG AGTCATTCAA TCCATCTCCA ACCATCAAAA TAGTGTGACC 60

TGCTTTCTGC AGTTTCTCTA CTAACTCAAA TTTCCCATCA GGTTTCAAGT CTGTATAGAC 120

CTGATCAAAG GGCAAATCTT TGACTAATTC CTCTGTCCTA ATCAAGGTGT CTCCTGTTGC 180

CAGAATCAAT TTTTyCCCCT GTGCCTTAAG TTTATCCAAG GCTGTTTTTG CTTCTTTTCT 240

CAAAGGAGTA TGAATGCAGA ACATTCCAAT CAATTCATTT TGATAAGCCA AGAATAAGAG 300

ATTGTAGTGA CTCTTGTACT CTTCAATTAA AGCATTTTGT TCTGAACTGA TATGAATCTG 360

CTCATCCTGC ATCAAGACAT AATTCCCAAT AAGAACTGGT TGGCCATCTA TATGAGATTT 420

GATCCCCTTG CTTGCGATAT ATTGGAGTTT CCCATGCATT TCCTCATGTT CAATTCCCTC 480

TATCTCAGCT TGCTTGACGA TGGCATTAGC AATAGGATGA TAAATGTGTT CCTCAAGACA 540

GGCACTGATT CTGAGAATAT CTTCCTCACT ATAGTCTCCA AAAGGTAACA CCTTTTCAAC 600

TATAGGATAA CTAGTTGTGA TTGTTCCTGT CTTATCAAAC AAGAAAGTAT CAACTTCCAG 660

ATATTTCTCC AGAACATCTC CATCCTTAAT CACCATTTCA CGGTTCAACC CTTCCTTGAT 720

AACTGTCAAA TAAGCTACAG GAGTAGAGAT TTTCAAAGCG CAGGAGAAAT CGACCAATAG 780

GAAAGAAATA GCCTTAGAAA AAGAACCTGT CAATAGGTAA GTCAGCCCAG CCCCCAAGAA 840

ATTATATTTG ACGACTTTAT CCGCCATCTT GATGAAATAG CGTTGTTTCG TTTTCTTGTT 900

TTCTTCAGAT TTCTTCATCA ACTCAATCAG CTGTAAAATA CGGCTGTTCA TCTGATTATC 960

TGTTACACGA ATGCGTAACT CTCCAGTTTC TAATACTGTA TTTGCACAAA CCAAATCAGA 1020

CTCTCTTTTT TCAACTGGAA AACTCTCTCC TGTCAAGGAA CTTTCGTTGA CCATACCTAA 1080

ACCTGAAACT ACTTGTCCAT CAAACAGAAT TTCATTTCCT TGAGATAAGA TCAAGACATC 1140

TCCTATTTGA ACATCGGAAC TCTTGATACT AACAACCGTA TCGCCCTGTA CTAGGAATAC 1200

ATCGCTCTCT TTTGCAAGAA GACTCTGTTC TAAATCTGTT GCAGTTTTTT TCAAGGACCA 1260

CTGATCTAAA TGATTCCCCA AATCAAGCAT AAACATGATA TTGCTAGCTG TCTTGGATTG 1320

GTTCATAAAC AAAGACAATA AAATAGCCGA ACAGTCCAAG ACTTCCATCG TTAGTyCCTT 1380

ACGCGCTAGT GTTTGATAGG CTTCTCTAAT ATAACCCAAA GCCTGATAAC AAGTCCATAT 1440

ATAGCGAATA GGATACGGCA CAAAACTACG AAAAAGTACA CGCTTAACCG CTGCACCTGA 1500

AACAATAGAA TAAGCACTCT CTTCTCTACG AATGGGAAGA GTCATCAACT CAGAAACTTT 1560

CCCTTTATCA ATTCTTTTTA AAAAGGCTTC TGCATTATCT AATACAGAAA AGCCTTCTTT 1620

TATGCGTAGA GTAAAGTGCT GTTGATCCAT GTAAAACTGG ATAGACTCAA TCCCCTTTTC 1680

ATCTCTCGCC AAGGAACGAA GATAGTCTTG AATATCCAAG GTAAGTGAAA AAGAAGATGA 1740

TAGTCGGATA TGTTGGTATC CTCTATGTAG CACTTTAAAA GACATATTAT TCACCTATAA 1800

GGCTATCTAA TTGCTCTTCT TTTTTCTCTT GCTCGTACAA ATATTTGGCA TCTTGCAAGA 1860

CATCGTCTCC ATGTTGCTTC ACAACAGAAA CAGATGCATC TAGCTCGTCT TTCAACTTGT 1920

AAGCCTTAGC CAAAGCTTTA GAATAACCTT TTTTAGCTTC CTTACTTGCT AAGATTTTCA 1980

AACCAAGGGT ACCAAATGCG ACACCACCCA AAAATAATGA AGATTTTTTC GCAACTTTTG 2040

CAACGGTTAA TACTTCTTTT AACATAGGG 2069

(2) INFORMATION FOR SEQ ID NO: 175:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4597 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

AAATCTTGCG CAATAAAGCT CATCTCCATC TCCCGATTGA AACAGTCACT CCCCGGACTG 60

TTTCAACGTC CCAAGACATA ATCTTAGGCA GATTTCTAAA ATTACACTCA AAGTGGAAGT 120

CATTGAGCTT TCGAATGACA GTTGAAGTTG AAATGGCCAG CTGATGGGCA ATATCGGTCA 180

TAGAAATCTT TTCAATTAAC TTTTGCGCAA TCTTTTGGTT GATAATACGA GGAATTTGGT 240

GATTTTTCTT GACGATAGAA GTTTCAGCGA CCATCATTTT CAAGCAATGA TAGCACTTAA 300

AACGACGTTT TCTAAGGAGA ATTCTAGTAG GCATACCAGT CGTTTCAAGG TAAGGAATTT 360

TATAGGGTCT TTAATGTCTA GTAATTTTGT GATAAAATGT AATTGTTCCA TATGATTCTT 420

TCTAATGAGT TGTTTTGTCG CTTTTCATTA TAGATCTTAT GGGACTTTTT TTCTACCCAA 480

AATAGGCTCC ATAATATCCA TAGGGAATTT ACCCACTACA AATATTATAG AGCCCAAAGT 540

TTTAGGTCGC TTGATAATAT GCGTTTTTTG AATTTTATAG ACTGCTCGTT TAAACTCTAT 600

TTACTTCGTA CCTTCTGGAG CGAGACGGAA TATTAGTCAC ATACAAAATG AGTACTATTA 660

GGATTTTATT TTCATGTACA ATTTCAGCCA GTCTTGTTAT AATCAGCCTA TAGGAATCAA 720

GGAGGTGACT CTTATGGCTG TTTTTGTGTC TTTGGATGGA ATTGTGGTAG AAGTCCTTGA 780

TGTCTTTTCT TCTTTTAATG GGGATAGTGA GTTTTTCTTG TGTATAGCAT TTTGAATCTG 840

GAATAGGACG CCATGACTGC TAAAAGATTT CTATAAATTA ATTTGATTTT CCTAATCAAT 900

TTGTTCATAT CTTATTTCAT TCCACTATAA ACGTCTTAAA GACAAGAGTC AGTTTGTTAT 960

GGAACGCTCT CAGTTCGAGG AGATGTTCCA ACTTCAAAGT AGTCGCTTGA CGACGCAAGA 1020

AAAATTACAA TTGTTTACCT CTGTGTTTGC TGGCCGTTAT GATGTTTATG CTAAGAATTT 1080

TATCAATGAA CAAGGGAAAA TTCAGTATTT TCCTTCCTAT GATTATGGTT GGAAGCAGTT 1140

GCCACCTGAA AAACGGAGTT TCCAGACATT GACGAACTCC GTTTTGAAAT CTCATTTTCG 1200

TGGGGAGGCA GCTATCGGTA TCTTTCCTAT GCACTTAGAT GATAGCTGTT ATTTTTTGGT 1260

ACTGGATTTG GATGAAGGAG ATTGGAAAGA AGCTGGTTTA ACCATTCGAA GAATAGCCAG 1320

GGAACGCCAG ATGGAAGCCC ATTTAGAGAT TTCTCGTTCG GGTCACGGAC TCCATATTTG 1380

GTTCTTCTTT GAGGAAGCGA TTCCGAGTCG AGAGGCTCGC TTGTTTGGAA AGAAACTGAT 1440

AGAACTGGCA ATGCAGGAAA GTATGCAACT GTCCTTTGAT TCTTTTGATC GCATGTTTCC 1500

AAATCAGGAT GTCCTTCCTA AGGGGGGATT TGGAAATTTG ATTGCCTTGC CTTTTCAAGG 1560

AGAAGCTTAC CATCAAGGGC GAACGGTCTT TGTGGATGAA CAGTTTCAGC CTTATGAAGA 1620

CCAATGGAGG TATCTACAAG AAATTCAGAG GATTTCAACT GCTAAAGTGG CACTGTTAAT 1680

CCAAGAGGAG TTAGGCAAGC AAGAATTGGA TAAGGAGTTG AAGGTCGTTT TATCCAATAT 1740

GATCCAACTT GAAAAATCGT CTGTGACATC CAAGGCACTT TTTTCTTGAA AAATATGGCT 1800

TCCTTTTCTA ATCCCGAATT TTATAGTAGA TTGAAACTAG AATAGTACAC CTCTGCTTCT 1860

AAAACATTGT TAGAAATCGA TTTGACTTTC CTGATCGATT TGTCCTGTTA TTATTTCATT 1920

TTACTATATT TAAAGCAGGC TATGCGACAG CCAACCTATC AAATTCCTGA GAGAATGTAT 1980

TTATTTGGAG AATCCGATCA TTATTTATGG TTGCCAAGAG GTTTGCTGTA TCCATTGCAA 2040

GATAAATTTA AGCAGGTATC TGTGGAAGAT AGGAGAAAGG TACAAAGGTC TATTAGCGTG 2100

GAATTTAAGG GAGAACTCAC TTTTGAGCAA GAGTTAGCCC TGTCAGATAT GACTTCTAAA 2160

GAAAATGGTT TACTTCATGC GGAGACTGGT TTTGGGAAGA CCGTTTTAGG TGCTGCTCTT 2220

ATCTCTGAAC GGAAAACAAA AACAATTATT CTAGTCCATA ATAAGCAACT CTTAGACCAA 2280

TGGCTAGATC GCTTAAACTG CTTTTTGACT TTCGAAGAGG AGGAGGCTAT CCGTTATACG 2340

GCATCAGGTC GTGAAAAGGT AATCGGCTAT GTTGGGCAGT ACGGTGGGAC TAAGAAATGG 2400

CTGAGTAAAC TGGTTGATGT CGTTATGATT CAATCTCTAT TTAAGTTGGA AAATAGTCAA 2460

AGTCTTTTGG ATGAGTATGA GATGATGATT GTGGATGAGT GTCATCATGT CTCTGCCTTG 2520

ATGTTTGAAA AAGTTGTTGC TCAGTTTAGA GGGAAGTATC TTTACGGTTT GACGGCTACG 2580

CCTGAGCGTA AGAATGGTCA TGAGCCTATT GTTTTTCAGA GAATTGGTGA GATACTCCAT 2640

ACTGCTGATA AGAGGGAAAC GGATTTTAAA CGGCAATTGC AATTAAGATT CACTTCTTTT 2700

GGTCATTTGG AAATTGAAAA GACCAAAGCA AGTAATTTTA TACAGCTTAG TGATTGGATT 2760

GCTACTGACT CAGTGAGGAA TCAGATGATT CTCAAGGATA TTCTAGCCCA AGTGGCAGAA 2820

GGACGGAATA TCTTGGTTTT AGTTAATCGA ATTCAACAGA TAGATGTCTT TGAAAAATTA 2880

TTGAAAGAGA AAGAGGTTGA TGACTGTTAC ATTATTAGCG GAAAAACCAA AGTCCGAGAG 2940

AGAACGAGTT TACTGGAGAC GTTAGAACAG TTAGATAAAG GGTTTGTTTT GTTGTCTACT 3000

GGAAAATACA TTGGCGAAGG TTTTGACTTA CCTCAGTTGG ACACGCTTAT CTTGGCAGCA 3060

CCCTTTTCTT GGAAAAATAA TTTGATTCAG TATGCAGGTC GGATTCATAG AAACTACAAG 3120

GATAAGTCTT TGGTGCGTAT TTTCGATTAT GTGGATATTC ATGTTCCTTA TTTAGAAAAG 3180

ATGTTTCAGA AACGACAAGT AGCTTATCGA AAGATGGATT ATCGTGTCAT CGAGGGTGAG 3240

GAGAAACAAT TCGTTTATGT TGATAGTAGA TATGAGAAGG TGTTGAGAGA GGACTTAGCA 3300

GGGGAAAGAC AGGAATGTCT GCTTATTTTA CCTTATGTGC ACCAGACAAA ACTGATGAAT 3360

TTTCTAAAAG AATTTAGGAT TAGTCAAATT GAGATATGTA TACCAGAGAC GGTTGCAAAT 3420

AAAGCATGGC TAGACCAGTT GAAGAGCCAG AAAATTAAAG TGTCTTTTAC TCAATCAAAA 3480

ATAGTAACGC CTATTCTTTT GGTGAATAAG ACTATTGTTT GGTATGGTGC AATGCCATTA 3540

TTAGGGAAGG TAGATGAGAT GACCATATTA CGTTTGGAAT CAGCTAGTAT AGTTTCTGAA 3600

CTAGTGGCAG GTTTACGATA GAGAAAATTT TTAAAAATTT CTATGTATGA TTTTCATTTC 3660

TTTAGTGAGA CTGTTGCCAT TATCACATTC GAATCACACA AAATAAAAAA ATTTTTATAA 3720

GTACTTGACA AATAGATTGA AATATCATAA AATAAAAACG GTTACAGAGT TATTAATTAT 3780

TTAAGCTTCA TGTCACCATT AAAAATTGAA ATAAAAGGAT GTTATCACTA ATACAAGTGA 3840

GCAGGAACCT ATTTAATCAC ATCAGAAGAA GTTTCTTGAT GTTTTTAAGT AGGTTCCTTT 3900

TATTTTAAAA GGGAAATTTT ATGATCATAA AACGAATACT AAACCACAAT GCCGTAATTG 3960

CGCAAAGTAA AAAAGATATC GATATTCTTC TTTTTGGAAG GGGAATAGCT TTTGGAAGAA 4020

AAACTGGAGA TAAAGTAAAT CCAATTGATA TTGAGAAAAG TTTTTTTCTC AAAAATAGAG 4080

ATAATATGAC CCGTTTTACA GAGATGTTTA TTAACGTTCC TTTGGAGTTG GTGTACATCA 4140

CCGAAAAAAT AATTAACCTA GGTAAAATAA CATTGGGTAA TAATTTTGAT GAAATTATCT 4200

ATATTAATTT AACGGATCAT ATTTCTTCGA GCATAGAACG TTATAAAGAA GGGATTATTA 4260

TTTCGAATCC CCTACGCTGG GAAATATCGA AATATTATAA AGAAGAATTT GAACTTGGGA 4320

AAAGGGCTTT ACAAATAATA AAAAAAGAGT TAGGTATTGA ACTTCCAATT GACGAAGCTG 4380

CATTCATAGC GCTACATTTT GTTAATGCTA ATTTAGAAAA TAATTTTCAA GAGTCGTATA 4440

AAATCACTGA AATAATTATG GGAATTGAGA AAATCATTCA AGATTTCTAT TGTACTGAGT 4500

TTAACCAAGA TTCTATTGAT TATTATAGAT TCATAACTCA TATGAAATTA TTTGCCCATC 4560

GCTTGGTTGA GAATACAACT TATTGTGACG ATGATGA 4597

(2) INFORMATION FOR SEQ ID NO: 176:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3984 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

CGGCTTATTT ACTACTTGTT CCATCATATA TGGAATATGC ATGAAACCTG CTCTCATATT 60

AGGGAATTTT TTATCCACTA AATAAAGAGC TTGGTACATC AAATGATTGC AAACAAAGGT 120

TCCTGCACTA TTGGATACAA CTGCCGGAAG TCCCTGTTTT TTGATAGCTT GTACCATCGC 180

TTTGATAGGT AAACTACTAA AATAGGCCGA TGCTCCATCA ATACGAATCG GTGTATCAAT 240

TGGTTGATTG CCTTCGTTAT CAGGTATGCG AGCATCATCT TGATTAATAG CCACTCGTTC 300

AGGTGTTAAG CCGGTCCTGC CGCCTGCTTG TCCAATACAA AGTACAGCAT CTGGTTGATA 360

TCGTAATATT TCTGCCTCTA AAACTTCTGA CGACTTATAA AAAACCGTTG GAATTTCTAC 420

CCAGCGAACT TCAGCCCCAT TAATCTCAGA TGGTAATAAT TTTACAGCCT CCAAAGCTGG 480

ATTAATCTTT TCACCTCCAA AAGGATTAAA ACCTGTAACC AATATTTTCA TTTTATTTTC 540

CTTTACTAAA ATGCGAGAAA GTACATTAAG AATATGTGAA TAACAATCAT TACTAGAGCA 600

ACACCTGCTT GAGCCTTTAT AACGCCATTC TGATCTTTCA TATCCATCAA TGCTGCTGGT 660

AGAGCGTTAA AATTAGCAGC CATTGGGGTC AATAAGGTCC CACAATAACC TGCTGTCATG 720

GCAAGAGCAC CAGCCACAAT TGGATTAGCT CCCAGAGCAA ATACAAAGGG AACTCCAACA 780

CCTGCTGTAA TAACGGTGAA TGCTGCAAAA GCATTTCCCA TAATCATTGT GAATAGAACC 840

ATTCCAAGAA CATAGGCCAA AACTCCTATA AAGCGACTAT CTGAAGGAAC AATACCGCTA 900

ATCAGATGAG AGATAACATC ACCAACACCT GCTACAGTAA AAATAGCCCC CAAAGCCCCT 960

AATAATTGAG GAACAATCCC ACTTGTTGAA ACTTGCTGAG TCATTCGATT ATTTTCTGAT 1020

AACAGACTCT TAGGGTGACT ATTGGTAATC ACAAGAACAG AAATTGTAGC AAACAAGGCG 1080

GCAAGGCTAA TCGAAATCTT GCTAAATTCT GGAATCATTT GCGCTAAGAC CAACGCAAGT 1140

ATTGCCATCA GCATAACTGG AATAAAAATT TTATTTTTCA ACCTGTTAGA TTCAATATTG 1200

GCTTTCATTT CATCTAAGGA TGGCAAGGTT CCGATACGGA CTTGCTTAAA CAATGTTAAC 1260

AGCGATAATA GGATTACAAT AATACCAATA CTCATATTTG GCATATAGGA ACCACCTATA 1320

AACGTAATAG ACAATAGAGT CCAAAATGCA GATGTCCCAA GTCGAACTGG GTTTGTTTTA 1380

TCTTTATAAC TACAATAGGC TGTATGGAGA AATTGACAAC CAATCACAAT ATAGGTCAAC 1440

TCTAATAGTT GCTTTGCCAA CTCTGTCATT TTTGTTCTCC TCCCCTAGTC TTTTTTGATA 1500

TCAATTTTTT ATCAAATAAA TAATTATAAA TCCCCACTAC AATAAGTGTT ATAACAGCAA 1560

CAATAATAGA TGTAGAAGCA ATCCCTGCAT AATTGCTTTC ATAGCCTAAC TGATCTAATG 1620

TTCCCCCTAT CAAGAGGACT CCCCCAGCAC CTACAAACGT ATTTTGAGCA AAGAAATTTC 1680

CAAAATTTTC ATTCGCAGCC GCACGCGCTT TTATTGTCTC ATCTTCAACC TCTGTTAACT 1740

TTCTACCTAA TTGAGACTCT GCAGCTGCTT CTCCCATAGG TTGAACCAAA GGTCTGACAA 1800

ACTGAGGGTG TCCTCCTAGA CGAATTGAAA AGAAACCAGC TAACTCTCGA ATAAAGAAAT 1860

AAACTGTATA GAAGTTTCCA ACTGTCAGAC CTTTAATCTT TCGAATCAAA TCGATTGATC 1920

GTTGCTTGAG TCCAAAGGTT TCTGACAGCC CCACAAGAGG CAAGGTAACC ATAAAAATCG 1980

TGAGCACTCG CTGATTGCTA AATTCTTTTC CCAAAATCTC CAAAAATTCA ACGAGAGAAA 2040

CACCTGAAAC TAAAGCTGTA ACCAAACCAG CTAAGACTAC TGTTGCAATT GTATCAAATT 2100

TTAAAATAAA ACCCACAACA ATGATTGCTA TTCCTATTAA TCTAATCCAC TCCATATCAA 2160

ACTCCTTTAT ATTCAAAATG ACAGTATTTT TAAAATTTTA TCAAGATCAA TACCATTCCT 2220

TATTTAATGT GTTTTTCTAG TTCTTTTTGG TATTTGCTAT TGGATTCCAA TTTTTCTTTT 2280

TGCCATTTTT TAAAAACCTC GTTATATTCT TTTGTTGTAA CAATATCTTT TTGCAATTTC 2340

ATTCCTTTAA AGATATATGG ATCCCCCTTA ATACCAACTT GTGAGTATGG TTTTGAGAAT 2400

GGTACTACGT TACTTACAAC TGGAGAACCA CCAGATGAAG CTGTTGGCAT CAATAATGAA 2460

CTATCTGTCG ACCAAGCTTG AGCTTTGGCA TATTTTTCAT ATCTTTTCTC TAGGTCAGTG 2520

GTCTCAGAAA CAGCATCTTC TAACAATTTC TTATATTTAT CCAAACCAGG TTTAGCTACA 2580

ACATCCTTAT CTTTTCCTTT CGTAATACCA AGGTGTTTCA TGGCAGAACC AGATTTTGGA 2640

TCTATAATAT TCAAGTGAGA CGCTGGATCT TGATAGCTTG GAGCCCATCC TGTACTGTTC 2700

AAATCATAGT CTTTTTGAGA AGGAGCAACA TTGCCGTATT TATCATTTTC CATCAAACCA 2760

TCAATAACAT TTCCAATAAC GTCTGTCCTC GATGTTCGAG TCGCTATACT GTAGCCCAAT 2820

GATGCTGGAT CTACTGCATA GACATAAGAA AATGTTGTCG GTGCATCTGC TTCTTTATCA 2880

GTTTTTCCAC AAGCCACTAA AATAGCTGAC GTGCTCAGGA CCACTCCTGC TGTTAAGAGC 2940

CACTTTTTCT ATTTCATAAA GAATCTCCTT TGGTTTATTT TAATCTACTT TTACAATCCA 3000

ACCTTCTGGC GCTTCAATAT CGCCAAACTG AATACCCGTC AATTCATTAT ATAATTTACG 3060

CGTCACAGGA CCTACTTCTG TTTCACTATA GAATACATGG AAATCATCAC CATGTTGAAT 3120

ACCTCCAATT GGAGAAATAA CCGCTGCTGT ACCACAGGCA CCTGCCTCTA CAAAACGGTC 3180

AAGATTATCA ATTGGAACAT CACCCTCAAT AGGAGTTAAT CCCAAGCGAT GTTCTGCCAA 3240

ATAAAGCAAG GAATACTTGG TAATAGATGG CAAGATAGAT GGACTCAATG GTGTTACAAA 3300

TTCATTATCA GCTGTAATTC CAAAGAAGTT AGCTGATCCG ACTTCTTCAA TCTTTGTATG 3360

AGTTGATGGG TCCAGATAGA TAACATCTGA GAAATGACGT GACTTGGCCA TTTTTCCTGG 3420

TAAGAGACTT GCAGCATAGT TTCCACCAAC CTTAGCCGCA CCTGTACCAT TTGGTGCTGC 3480

ACGGTCGTAC TCATCCTGAA TCAAGAAGTT GGTTGGGACC AAACCACCTT TAAAGTAATT 3540

TCCAACTGGC ATAGCAAAGA TGGTGAAAAT GTACTCTTCT GCCGGTTTTA CCCCGATAAT 3600

ATCTCCGACA CCAATCAAAA GAGGGCGAAG ATATAAGGTT CCACCTGTTC CGTATGGTGG 3660

TACGTATTCT TCATTCGCAC GGACAACTGC TTTACAAGCT TCTACAAACA TGTCTGTCGG 3720

AACTTGTGGC ATCAAGAGAC GGTCACATGT ACGTTGCAGA CGTTTAGCAT TTTCATCAGG 3780

ACGGAACAGT TGAACACTGC CATCCTTAGT ACGATAAGCT TTCAAACCTT CAAATGCTTG 3840

TTGTCCATAG TGAAGACTTG GAGAAGACTC TGAAATATGC AAAGTTGCAT CCTCTGTAAG 3900

CTCTCCTTGA TCCCATTGTC CATTTTTGAA ATGAGCAAGA TAGCGATAAG GTAATTTCAT 3960

ATAGGAAAAA CCGAGGTTTT CCGG 3984

(2) INFORMATION FOR SEQ ID NO: 177:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8703 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

TATCTAATTA TTGGTTTTTA TCGCTGACCT TGGCTATTGT TGGGGTTGTT TTACCCTTGT 60

TGCCTACAAC ACCTTTCCTT TTGTTGTCTA TTGCTTGTTT CTCCAGAAGT TCCAAGCGAT 120

TCGAAGATTG GCTTTATCAT ACCAAGCTCT ATCAAGCATA TGTAGCTGAT TTTCGTGAGA 180

CCAAGTCTAT TGCGCGTGAA CGAAAGAAAA AAATCATCGT CTCTATCTAC GTCTTGATGG 240

GAATTTCTAT TTATTTTGCA CCTCTTTTAC CAGTCAAAAT CGGTCTGGGT GCTTTGACCA 300

TCTTTATTAC TTATTATCTC TTCAAGGTCA TTCCAGACAA AGAATAGTTA AAACAGTAGT 360

TATTTGCCTT GATAAAATTG AAAGCATATT CATAACAATA TGATATAATA AAATTGAAGT 420

AATATTCAAG GAGAATCAAA TGATTTACGA ATTTTGTGCT GAAAATGTGA CTTTACTTGA 480

AAAAGCGATG CAGGCTGGAG CTCGTCGGAT TGAACTCTGT GATAATCTAG CAGTTGGTGG 540

GACAACACCC AGCTATGGAG TGACTAAGGC AGCGGTTGAA CTGGCAGCTA ACTACGATAC 600

AACCATCATG ACCATGATTC GGCCACGTGG TGGTGACTTT GTCTATAATG ACCTAGAAAT 660

TGCTATCATG CTAGAAGACA TTCGTTTGAC TGCTCAGGCT GGAAGTCAAG GGGTTGTATT 720

TGGAGCTTTA ACTGCTGATA AAAAGTTGGA TAAGCCTAAT CTGGAAAAGT TAATTGCTGC 780

ATCAAAAGGA ATGGAAATTG TCTTTCACAT GGCCTTTGAT GAACTAAGTG ATGAAGATCA 840

AGCGGAAGCT ATTGACTGGC TCAGTCAAGC CGGTGTCACT CGTATCCTAA CTCGTGCTGG 900

TGTGTCTGGC GACTCCTTAG AAAAACGTTT TGTTCACTAT CACAGAATTT TGGAGTACGC 960

TAAAGGTAAA ATTGAAATTC TACCAGGTGG GGGGATTGAC CTTGAAAACC GTCAAACCTT 1020

TATCGACCAG GTGGGGGTAA CACAATTGCA TGGTACTAAG GTTGTTTTTT AAAAAATAGA 1080

AAGGAACTGC TAGCTTTGGG TAGCAGTTTT CACTTATGTT TGAAATTTTT AAATCCTATC 1140

AATTTAATCA AGAAAAGGCT CATGATTATG GTTTTATAGA AAATAGCGAA GTCTGGACAT 1200

ATAGTTGCCA GATTTTGCAA GGTGACTTTG TCATGACTGT GTCCATCACT GCTGATAATG 1260

TGAACTTTCA AGTCTTTGAC CAAGAGACTG GTGACCTCTA TCCTCACGTT TATATGGAAA 1320

GCATGAGGGG AAGTTTTGTC GGAAATGTCC GTGAGGCTTG TCTGGAGATT CTTTACCAGA 1380

TTCGGAAGGC TTGTTTTGAT GTGCAAGATT TTATCTGTCA TCAGACTAAG CGTATCATGA 1440

CTCAAGTTCA GGAAAAGTAT GGAAACCAGT TGGAGTATCT GTGGGAAAAA TCGCCTGATA 1500

CAGCTGTATT GCGCCATGAA GGCAATCAAA AGTGGTATGC CGTCTTGATG AAAATCTCTT 1560

GGAATAAGCT GGAAAAGGGC AGAGAAGGAC AAGTGGAAGC AGTCAACCTC AAGCATGACC 1620

AAGTAGCTAA TTTGCTTTCA CAAAAGGGGA TTTATCCAGC CTTCCATATG AGCAAGCGCT 1680

ACTGGATTAG TGTGTCCCTT GATGATACTT TATCAGATGA AGAAGTACTG GAATTGATAG 1740

AAAAAAGTTG GAACTTAACC TCTAAAAAAT GAAATATTTT AATAATTTTC ATGAACTTTC 1800

AATTAGCTAA ATATTCTTTA CTGAAGAGAT TTTTAGAAAA TATAGGATTT ACCACACTAG 1860

AGGAATATGG TGCCATCTTC AAATACCTGA TTGAGAATGT CAAGACGGAT CGTCAGATCA 1920

TCTATTCGCC TCACTGTCAT GATGACCTCG GAATGGCAGT GGCAAATAGC CTTGCTGCTG 1980

TCAAGAATGG TGCAGGACGT GTTGAAGGGA CTATCAATGG TATTAGGGAG CGAGCTGAAA 2040

ATGCTGCTTT GGAAGAAATT GCAGTGGCTC TCAATATTCG CCAAGATTAC TACCAAGTAG 2100

AAACCAGTAT TGTCCTAAAT GAGACCATCA ATACGTCAGA AATGGTTTCT CGCTTCTCTG 2160

GTATTCCAGT TCCTAAAAAC AAAGCCGTCG TTGGTGGCAA TACCTTCTCC CACGAATCTG 2220

GTATTCACCA AGATGGAGTC CTTAAAAATC CTCTCACTTA TGAGATCATC ACACCTGAAT 2280

TGGTTGGTGT TAAGATTCTG CTTGGAAAAT TATCTGGTCG CCATGCTTTT GTTGAGAAAC 2340

TGAGAGAATT GGCCCTAGAT TTTACAGAAG AGGATATCAA ACCACTCTTT GCTAAGTTCA 2400

AGGCACTGGT CGATAAGAAG CAAGAAATCA CAGATGCAGA TATTCGAGCT TTGGTAGCTG 2460

GAACCATGGT TGAAAATCCA GAAGGCTTCC ACTTTGATGA TTTACAACTT CAAACTCATG 2520

CAGATAATGA CATTGAAGCG CTCGTTAGCC TAGCCAATAT GGATGGTGAG AAAGTCGAAT 2580

TTAATGCGAC AGGGCAAGGT TCCGTTGAAG CAATCTTTAA TGCTATCGAT AAGTTCTTTA 2640

ACCAATCTGT TCGTTTGGTG TCCTACACTA TCGATGCGGT AACAGATGGA ATCGATACCC 2700

AGGATCGGGT TTTGGTCACT GTTGAAAACA GAGATACAGA AACCATCTTT AATGCAGCAG 2760

GGCTTGATTT TGATGTGTTG AAGGCTTCTG CTATTGTCTA TATAAACGCT AATACCTTTG 2820

TTCAAAAAGA GAATGCAGGT GAGATGGGAC GCAGTGTTTC TTACCACGAT ATGCCTAGTG 2880

TGTAAAGGAG AAGGCTATGG CAAAGAAAAT AGTAGCTCTA GCAGGAGACG GAATTGGCCC 2940

AGAAATCATG GAGGTTGGTT TAGAAGTTCT GGAGGCTCTA GCTGAAAAAA CAGGTTTTGA 3000

CTATGAGATT GACAGACGAC CGTTCGGAGG TGCAGATATT GATGCAGCAT GACCTCCCTT 3060

ACCTGATGAA ACCCTTAAGG CAAGTAGGGA AGCAGATGCT ATCCTACTAG TAGCTATCGG 3120

TAGTCCTCAG TATGATGGAG CAGTGGTTCG CCCTGAACAA GGCCTGATGG CTCTCCGTAA 3180

GGAACTCAAT CTTTACGCTA ATATTCGTCC TGTAAAAATC TTTGACAGTC TCAAGCATTT 3240

GTCACCACTC AAACTGGAAC GAATTGCTGG TGTAGACTTT GTCGTGGTGC GTGAATTGAC 3300

AGGCGGGATT TACTTTGGAT ATCATATTCT TGAAGAGCGC AATGCGCGTG ATATCAACGA 3360

CTATAGCTAT GAGGAAGTGG AGCGGATTAT TCGCAAAGCC TTTGAAATTG CAAGAAATCG 3420

CAGAAAAATC GTTACTAGTA TCGATAAGCA AAATGTTCTA GCGACCTCAA AACTCTGGCG 3480

GAAAGTAGCT GAGGAAGTCG CACAGGATTT CCCAGATGTA ACCTTGGAAC ATCAGCTGGT 3540

AGACTCAGCT GCTATGCTTA TGATTACCAA TCCTGCTAAG TTTGATGTTA TTGTAACGGA 3600

GAATCTTTTT GGAGATATTT TATCTGATGA ATCAAGCGTC TTATCTGGTA CACTTGGGGT 3660

TATGCCATCA GCCAGTCATT CTGAAAATGG ACCAAGTCTC TATGAACCTA TTCACGGTTC 3720

AGCACCTGAT ATTGCAGGTC AAGGAATTGC CAATCCTATT TCCATGATTT TATCAGTTTC 3780

CATGATGTTG AGAGATAGTT TCGGACGTTA TGAGGATGCA GAGCGTATCA AACGTGCTGT 3840

TGAGACAAGT CTGGCGGCAG GAATTTTAAC GAGAGATATA GGAGGTCAGG CTTCAACAAA 3900

GGAAATGACG GAAGCTATTA TTGCAAGGTT ATGAAGTTAG ACGAAAAAAT TACTCTAGTC 3960

CTTTTGATTT GGAATGTCAT CATTTTCTTG ATTTATGGTA TTGACAAATC TAAGGCAAGG 4020

AGAAGAGTTT GGCGCATCCC TGAGAAAATC TTACTTATTT TAGCCTTTAC TTTTGGTGGT 4080

TTTGGTGCCT GGCTAGCAGG AATCATCTTT CACCACAAGA CTCGAAAATG GTACTTTAAA 4140

ATAGTTTGGT TTCTTGGGAT GGTGACCACA CTAGTAGCCT TATATTTTAT TTGGAGGTAA 4200

TGGATGGCAG GGTCTTCGAG GGAATACGCT GCTTGGGCTC TAGCGGACTA TGGTTTTAAG 4260

GTCGTGATTG CAGGATCTTT CGGTGACATT CATTACAATA ATGAACTCAA TAATGGCATG 4320

TTGCCAATCG TTCAGCCTAG AGAGGTTAGA GAGAAACTAG CCCAGCTAAA ACCAACCGAC 4380

CAGGTAACTG TGGACTTGGA ACAACAAAAA ATCATCTCAC CAGTTGAAGA ATTCACCTTC 4440

GAGATAGATA GCGAGTGGAA ACATAAACTC CTAAATAGTT TGGATGATAT CGGTATTACC 4500

TTGCAGTATG AAGAGTTGAT TGCTGCTTAT GAAAAACAAC GACCAGCCTA CTGGCAGGAT 4560

TAGAAAAAAT AGAAAAGGAG ATATAGTAAA CTGAAATAAG ATGTAAACAA ATGAATTGGA 4620

GCTTAACATC CATTTCCAGC AATTTTTTAG AAACTACAGT GGACTATTCT GGATTCAACA 4680

CATTATAAAA TTATGACAAA ACACATTCAC AAGAAGGCTA CGACATTTTA AAAGGTGAGG 4740

GCGGATGTAT CGTTTGCCCT ACTAAAGTTG GTTACATTAT CATGACCAGT GACAAGGCAG 4800

GACTTGAGCG TAAGTTCGCA GCCAAAGAAC GTAAGCGTAA CAAACCAGGT GTTGTTCTCT 4860

GCGGTAGCAT GGATGAACTT TGCGCTTTAG CGCAACTCAA CCCAGAAATT GAAGCATTCT 4920

ACTAAAAACA TTGGGATGAA GATATTCTTC TTGGTTGTAT CCTTCCTTGG AAACCAGAAG 4980

CCTTTGAAAA ACTCAAAGCA TACGGGGATG GCCGTGAAGA ACTTATTACT GATGTACGTG 5040

GTACTAGCTG TTTTGTTATC AAGTTTGGAA AAGCAGGTGA ACAATTGGCT GCCAAGCTTT 5100

GGGAAGAAGG TAAAATGGTC TACGCCTCAT CTGCTTCAAT GACAAAACGA TTGAAACTCG 5160

CTATGAGCAA GGTGTAATGG TGTCTATGGT CGATAAGGAC GGCAAACTCA TCCCAGAACA 5220

AGGAGGAGCA CGTTCAACTT CACCAGCTCC AGTTGTGATC CGTAAAGGGC TTGACATTGA 5280

TAAAATCATG ATGCACCTGT CAGATACTTT TAACTCATGG GACTACCGTC AGGTTGAGTA 5340

TTATTAGGAT AGAGAAGAAG TCTAGTGTTA TGAGATATTA AAGCTCCTAA CACTGGGCTT 5400

TTGTTTAGAA TTTCTTTTCT TTTTCTATAG GATATGGTAT TCTATGTAGA AAATATATGT 5460

TAATAAGTAA TGCCAATATT TAAACATCAT TAGTAAAAGG AGTTAGATTG ATGAATAAAA 5520

GAAAAGTTAG TTTAGAAGAT TTTTATAAAT GGTATAGTCT AAATAAAGAA GAGTTATTAA 5580

ATAAGGCAAC TGTTGGTGAA AAGTTTAATG ATAAATTAAA AGAAGAGTTT CTCCAGGAAT 5640

GGCCTTTGGA TAGGATTTTA ACAATGTCAA TCGATGAATA TGTAATAGGA AAGGGACAGC 5700

AAAATAAGTC TTTATGCTAC GCTCTTGAGA AGGGAAAATA CAAAAATCTA TTTCTTGGAA 5760

TTTCTGGTGG CTCAGCTTCA AAATTTGGTA TTTATTGGAA TAAAAAAACA AACAAATATA 5820

AAGATCAAGC TAATAATGAG ATTTCAGAGT TGGATCAGCG ATTTTCAAAA TTAAAATCAG 5880

ATTTGTATGA AATTATCAAA GAAGGTATTC GTTTTAACTT TGAAAATCCT ATTTTTGATA 5940

TGAAAAGATC AACAAATGAA TTTATTGGTC GTTCTGCTAT GGTGACAAAA TTACTTTGTA 6000

TCTATACTGA GGGAGATCCT TTCTTTGGTG TAAATATTAA TAGTCAGAAA GAATTTTGGA 6060

ACCACTTTGT TTCTCAGACA AATCAAGGTG GACCTTATCT GCAAAATCAT AAAATAATTG 6120

AACTGGTGTC CAAAACTTAT CCTGAGTTGG AGCCATCGAA ATTAGGAACT ATGCTTTTTG 6180

AGTATTCTAA GCTTTTTATG GAAAATAAGG AAGACAATAG TACAATGGAT TCATCAAACA 6240

ATTTTCGTCA TCAATTAACT CAATCTCTAT TAAAGTCTCC AAACCTCATC CTCCGCGGTG 6300

CTCCTGGCAC GGGAAAAACT TATCTTGCTA AAGAAATTGC TAAAGAATTA ACGGATGGCA 6360

ACGAAGATCA AATCGGATTT GTACAATTTC ACCCATCATA TGATTATACG GATTTTGTAG 6420

AAGGTTTAAG ACCAGTATCA AATGGGGATG GAGCTATTGA GTTTAGGCTA CAGGACGGTA 6480

TTTTTAAAGA TTTTTGTCAG AAAGCAAAAG AAACCCAATT GATTGGAGGA CAAGATAATT 6540

TTGATGAGGC TTGGGATTCT TACTTAGAAT ATATAAATGT TGCTGAAGAA AAAGAATATA 6600

TAACAAAAAC ATCTTACTTA TCTGTTAATA GTAGACAAAA TTTGTCAGTA AATTATGATA 6660

GTGGTGTTCC AGGATGGTCA CTACCTAGCA AATATGTTTA CGAGTTGTAT AAAGATAAAA 6720

ATTATAATAA GCAAGAATAC TACAAAAGTG GTGGAAAAAC TGTCCTAGAA ACATTGAGAA 6780

AGAGATTTGG TTTGAAAGAC TATGTTTCCC CAACAGAAAT TGATACTGAT AAGAATTTTG 6840

TCTTCATCAT CGATGAGATC AATCGTGGGG AGATTTCTAA GATTTTTGGC GAACTCTTTT 6900

TCTCTATCGA CCCCGGCTAT CGTGGTGAAA AAGGAAGTGT TTCTACCCAA TATGCAAATC 6960

TACACGAAAC TGATGAAAAG TTCTATATCC CCGAAAATGT TTACATCATC GGAACTATGA 7020

ATGATATTGA TCGTTCAGTG GATACCTTTG ATTTTGCTAT GCGTCGTCGT TTTCGTTTTG 7080

TTGAAGTTAC TGTCGAGGGT CAAGCTGGCA TGTTGGATAA AGAGTTGAAT ATCCATGCAG 7140

AAGAAGCAAA AATTCGTCTA AGAAACTTGA ACGCTGCTAT CGAAAATATT CAGGAATTAA 7200

ACAGTCATTA TCATATTGGA CCAAGTTATT TTCTTAAGTT GAAGGATGTA GATTTTGACT 7260

ATGAATTACT CTGGTCTGAT TATATTAAGC CTCTCCTAGA AGACTACTTG CGAGGTTCTT 7320

ATGATGAGGT TGAAACTTTG GAaACTTTGA AAAAAGCATT TGATCTGACA AATAATGAGC 7380

AAAAAGATCA GGCAGTAGCT GATGACAATG AAGGCGATGA AAACGATGAT GCGGATTACT 7440

GATAATCAAC ACAAGATTAT TAAAGAAAAA TTTGTTGAAG AATATCCTAA ACTAAGCAAT 7500

CCTCTTTTAG ACAGAACCTT GGAAAGTCTA TCCCAAGATG AACGTATTTT CATTTTTCCA 7560

AATGATTwGA CTCATACTCC TGATTTGGAT AAGGACCAAA AGATTTTTGA AACAGTCAAT 7620

CAGAAAATCA AGACAGGGAA CGTGATTGGT TTTCTTGGAT ATGGTCAGGA AAGATTAACG 7680

ATTTCCTCAC GATTTTCTGA TGAGAGTAAT GACCACTTTT TGCATTATCT CTTAAACAAG 7740

GTTCTTCATA TCAATCTCAC TAGTTTAGAT GTTGCTTTGT CTCGTGAAGA GAGGCTTTAT 7800

CAACTTTTGG TGTATCTCTT TCCCAAGTAT CTACAAGCTG CTATTCGAAA AGGTCTTTAT 7860

AAGGAATATC ATCGATTTTC TCATAACGAC AGTCATGTTA AGGGAGTGAT TGATGTAAGA 7920

AACCATCTCA AGAAAAATCT TCCTTTCACG GGAAATATTG CCTACGCAAC GAGAGAGTTC 7980

ACCTATGATA ATCCCCTCAT GCAGTTGGTC CGTCACACTA TTGAATACAT TAAGAATCAG 8040

AAAAGCATTG GTCAAGGGGT ACTAGATAAT CTCTCAACTA GTCGTGAAAA CGTATCTGAA 8100

ATCGTGCGTG TAACGCCCTC TTATAAACTA GCTGATCGTG CTAAGATTAT TCGGGGAAAT 8160

CAATCTAAAC CTATACGTCA TGCATACTTT CACGAGTACA GAAACTTACA AGAACTTTGT 8220

CTGATGATCC TAAACCAAGA AAAGCACGGT TTAGGGTATC AAGATCAAAA AATCTATGGT 8280

ATTCTCTTTG ATGTTGCCTG GCTTTGGGAA GAGTATGTTT ACACCTTGTT GCCAAAAGGT 8340

TTTGTACATC CCAGAAATAA GGATAAGACG GATGGAATTT CAGTATTTTC TGTTGGGAAA 8400

CGAAAAGTAT ATCCAGATTT TTATGACAGA GAACGAAAGA TTGTTCTAGA TGCAAAATAT 8460

AAAAAACTGG AATTGACTGA AAAAGGAATC AACCGTGAGG ACTTATTCCA GCTGATTTCC 8520

TATTCTTATA TTTTAAAAGC TGAGAAGGCT GGACTGATTT TTCCTAGTAT GGAGCAGTCA 8580

GTAAATAGTG AAATAGGAAA AGTAGCTGGC TATGGAGCTC AATTGAAGAA GTGGTCTATT 8640

CGAATCCCTC AGAATGCCTC ATTCTATAGT ACATTTTGTA AAATGATGGA AAATTCAGAA 8700

GAG 8703

(2) INFORMATION FOR SEQ ID NO: 178:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4854 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

CATCACCAGT TTTAGATGGC TTTAACAGTG AAATTATTGC TTTTAATCTT TCTTGTTCGC 60

CTAATTTAGA ACAAGTACAA ACAATGTTGG AACAGGCATT CAAAGAGAAG CACTACGAGA 120

ATACGATTCT CCATAGTGAC CAAGGCTGGC AATATCAACA CGATTCTTAT CATCGGTTCC 180

TAGAGAGTAA GGGAATTCAA GCATCCATGT CACGTAAGGG CAACAGCCAA GACAACGGTA 240

GGATGGAATC TTTCTTTGGC ATTTTAAAAT CCGAAATGTT TTATGGCTAT GAGAAAACAT 300

TTAAATCACT TAACCAATTG GAACAAGCCA TTATAGACTA TATTGATTAT TACAACAATA 360

AGAAAATTAA GATAAAACTA AAAGGACTTA GTCCTGTGCA GTACAGAACT AAATCCTTTG 420

GATAAATTAT TTGTCTAACT GTTTGGGGGC AGTACACAAG AAAGCGCTTT AAAACCAGTA 480

GACCTTTTCA TAAGGTTCGC TTGATGTACC AAGATGAGGC TGGTTTCGGT AGAATCAGTA 540

AACTGGGATC TTGTTGGTCT CCAATAGGAG TAGGTCCACA TGTCCATAGT CACTATATAC 600

GAGAATTTCG CTATTGTTAT GGAGCTGTTG ATGCCCATAC AGGCGAATCA TTTTTCTTAA 660

TAGCTGGTGG ATGTAATACT GAGTGGATGA ACGCCTTTTT AGAAGAGCTT TCACAAGCTT 720

ATCCAGATGA TTATCTTTTA CTCGTTATGG ACAATGCTAT ATGGCATAAA TCAAGTACCT 780

TAAAGATTCC GACTAATATT GGTTTTACCT TTATTCCTCC ATACACACCA GAGATGAACC 840

CATTGAACAA GTGTGGAAAG AGATTCGTAA ACGTGGATTT AAGAATAAAG CCTTTCGAAC 900

TTTGGAAGAT GTCATGAATC AACTCCAAGA TGTCATACAA GGATTGGAGA AGGAGGTGAT 960

AAAGTCCATC GTTAATCGGA GATGGACTAG AATGCTTTTT GAAAACAGAT GAGTATAAAA 1020

TTGAATTGCT TATAAAAAAG CTCCATACAC TGGATGTGTA TAGAGCAATG GGGCTTTATT 1080

TGATATAGAG TTCTTGGTTT TTTAGGACAA TTTCTCGGAT ACTTGCAAAC TTTTTAAGTT 1140

TTTTGATTTC TTCTGGATGA GTGACGAGAG TGATAACATA ACCTTCCTTG CCCATACGAC 1200

CAGTACGGCC AGCACGGTGT GTGTAGGTTT CGCTATCTCT AGGAATATCA AAGTTTACGA 1260

CACATTCTAG GCTATCGATA TCAATTCCAC GAGCCAAAAG GTCAGTTGCA AGAAGCAGGG 1320

TTAGTTGGTT ATCTTTAAAC TTTTCTAAGA TGATTTTTCT AAATTTAACA TTAACATCAC 1380

TAGCGAGGGA AACAGCCAAT ATATCACGAT ACTGTAGTTT TTCCTCGGCA TTCCCAAGGT 1440

CTGACAGGCT ATTGAAGAAG ACTAGACCAC GGAAATCCTC TACATGAGCC AGTTTTCGTA 1500

GCATATCCAC TCGATGACGT TGGTCTACCT GCATGTAGAA ATGCTGGATA TTGTCCAATT 1560

TTTGATCAGA GAGATCAATA GTGCGTGTAT TCGGCACAAT CTTTTCTTGG TCAAACTTGG 1620

TCGTGGCACT CATGTAGACC AGTTGGTGGT CACGAGGTGC GTAGTGAGTG ATTTTTTCTA 1680

CAAAGTGAAT CTGAGAATCA TCTAGTAATT GGTCAAATTC ATCCAGGATG ATGGTTTCCA 1740

CATTCATCAT CTTGATTTTT TTAAGTTTAA TGAGTTCAAA GATACGGCCA GGAGTTCCAA 1800

TCAGAATTTC TGGCCCCTTT TTAAGACGTT CAATTTGTCG TTTCTGACTT GAACCTGAAA 1860

GGAAGAGTTG AGCAGTCAAT CCGATAGCTT CTGCCCACGT TTTACATACA TCAAAAATCT 1920

GTCCAGCAAG TTCCGTATTT GGTGCTAGAA TCAAGAGTTG TTGGGCTTTT TTCTTTTGTA 1980

GTCTGAGAAG ACTTGGTAGG AGATACGCTA GGGTCTTACC AGTTCCGGTT TGGCTCACTC 2040

CTAGGAGGTT TTCTCCAGCA AGAAGGGGCT CAAATAGTTG AGTTTGAATG GGGGTGAATT 2100

CTTGGAAACC GAGTTGGTCA CTCAGTTCTT GCCATTCAGT CGGTAGTTTG GTTTTCATTT 2160

TTCTGCCTCA AATCTAATGC CAGCAGTCTG GCGCATGGTA TATAGTAGCT CATGAACAGA 2220

GCCTGCATCA TACAGCCAAG TTTGGTAGAG ATTCAGATCT GGTTGCTGGA TCATGTGTGC 2280

AAATGCAGCG ACTTCCTCAG TCATCGTATG AGGAGCCTGT TGGATAGGAA GCTGGACTTG 2340

ATTTCCTTGG TGGTCGGTAA AAATAGCTGA GCGAATATGC TCAATCGTGT TGAGAGTCAA 2400

GGTTCCATCT GTTGTATAAA TCTCGCAAGG AAGATTGGAA GTGATGTTTT TTCCAGCCTT 2460

GATGTGAACT TGATAGTCTG GGTAGAAGAG GATACCATCT CCATTTAGGT CAATGCTATT 2520

GTCAAGCTGT TGAGCATGGT AAGTCGCGTC ATTGGCTTTT CCAAAAAGAC GAACAGCAGC 2580

ATAGAGGGGA TAAATCCCCA AATCCATGAG GGCTCCACCA GCAAAACGGT CTGAAAAGAC 2640

ATTTGGTGTT TGTCCAGCCA ACAAGTCAGG CATCTTGGAA GAGTATTTGG CATAGTTGAA 2700

ATCTGCTCCT AACACTTGCT TATCTGCTAA AAAGTTTTTG ATAGTAGTAA AGGCTTTCTC 2760

GTGGTAATTA CGAGCTGCTT CAAAGATAAA ACAGTTATTT TTTTCAGCTG TTTGAATCAA 2820

ATCAAACCAT TCTTGTGGTT GAGAGACAGC TGGCTTTTCG AGAATAACAT GTTTACCAGC 2880

AGACAAGGCA GCTTTTGCCT GAGCAAAATG TAAGGAGTTT GGACTGGCGA TATAGACTAA 2940

ATCAAAAGAA GATTTGAAGA AGACTTCTAA TTGATCGAAT AGTTGGATAT TCTGATAGCG 3000

AGAAGCAAAG GTTGCTGCAG TTTCTAGTTT TCTAGAATAG ATTGCGACCA GTTGGTATTC 3060

TCCACTGGTA TGGGCTGCTT CTATGAAATG ATGGCTGATA GCGCCAGTTC CGATGACACC 3120

TAATTTTAGC ATAAATACTC CTTTTCCGAT TTTAAATCCT TCTTTCATTA TAACATAGAT 3180

AGACGGGACT ATCCAACAGA GAGGAGAAAA TTTCAAATAA GCTATTAGCT TTCTTTTCCG 3240

AATAAATAGA TAGAAGCATA GAATCTAGCA AACCTAGATT TAAAAATGTG CTATAATAGA 3300

AGGAGGAAAA GGAGGATTCT CAGACATCTA GGTATCAGCC CAACTAATGA TTTGTCAATT 3360

TATCCGCGAT ATGCTGGACT TGCCAGCAAA AAATGTGACG ATTTTGGAGG GAAGTAACAT 3420

TCACGTCTTG CCTTCCATGC CCTACTCAGC GTAAGATTTC TATACTAGTA TAGACGTCTT 3480

GGCGGAGTTA GATAATGGAA TCCAAGTTAT CATCGAAATT CAGGTTCATC ATCAGAATTT 3540

TTTCATCAAT CGCCTATGGC CTTATCTGTG CAGTCAGGTT AATCAAAACC TAGAAAAAAT 3600

TCGCCAACGT GAAGGTGATA CCCACCAGAG CTACAAACAA ATCGCACTAG TATACGCTAT 3660

CGCAATTGTC GATAGTAATT ACTTCTCAGA TGACCTAGCT TTTCATAGTT TTATAGTAAA 3720

ATGAAATGAG AACAGGACAA ATCGATCAGG ACAGTCAAAT CGATTTCTAA CAATGTTTTA 3780

GAAGTATAGG TCTACTATTC TAGCTTCAAT CTACTAGAAA TTCCATAGAT AGAAAACTAC 3840

ATAATCTCTA CAGATACGGA TGTTGGAGTT GATGTAAGAT GCTTTGGCTT GCTAGAGGAA 3900

TTGTGGATTG CCAAATTGTA TCATTGAAAT TATTGCTCAA ATTTGTTATG ATATAAATAT 3960

GAATAAAAGT AGACTAGGAC GTGGCAGACA CGGGAAAACG AGACATGTAT TATTGGCTTT 4020

GATTGGTATT TTAGCAATTT CTATTTGCCT ATTAGGCGGA TTTATTGCTT TTAAGATCTA 4080

CCAGCAAAAA AGTTTTGAGC AAAAGATTGA ATCGCTCAAA AAAGAGAAAG ATGATCAATT 4140

GAGTGAGGGA AATCAGAAGG AGCATTTTCG TCAGGGGCAA GCCGAAGTGA TTGCCTATTA 4200

TCCTCTCCAA GGGGAGAAAG TGATTTCCTC TGTTAGGGAG CTGATAAATC AAGATGTTAA 4260

GGACAAGCTA GAAAGTAAGG ACAATCTTGT TTTCTACTAT ACAGAGCAAG AAGAGTCAGG 4320

TTTAAAGGGA GTCGTTAATC GTAATGTGAC CAAACAAATC TATGATTTAG TTGCTTTTAA 4380

GATTGAAGAG ACTGAAAAGA CCAGTCTAGG AAAGGTTCAC TTAACAGAAG ATGGGCAACC 4440

TTTTACACTT GACCAACTGT TTTCAGATGC TAGTAAGGCT AAGGAACAGC TGATAAAAGA 4500

GTTGACCTCC TTCATAGAGG ATAAAAAAAT AGAGCAAGAC CAGAGTGAGC AGATTGTAAA 4560

AAACTTCTCT GACCAAGACT TGTCTGCATG GAATTTTGAT TACAAGGATA GTCAGATTAT 4620

CCTTTATCCA AGTCCTGTGG TTGAAAATTT AGAAGAGATA GCCTTGCCAG TATCTGCTTT 4680

CTTTGATGTT ATCCAATCTT CGTACTTACT CGAAAAAGAT GCGGCCTTGT ACCAATCTTA 4740

CTTTGATAAG AAACATCAAA AAGTTGTCGC TCTAACCTTT GATGATGGTC CAAATCCAGC 4800

AACGACCCCG CAGGTATTAG AGACCCTAGC TAAATATGAT ATTACAAGCG GGGT 4854

(2) INFORMATION FOR SEQ ID NO: 179:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2186 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

TAAACAGGTG TTAGGTGCTC TAAACTATTA AAATTCTAAG GAAATAAGGC TACTTTTTCT 60

GGGTCTTGTT CATAGTAGGT GTGGTTCTTT TTTTCGAGTG TAGCCCATAG CTTTGAGCGC 120

ATAGTGGATG GTAGTTGGAT GACAGCCAAA TTCAGAAGCT ATTTCAGTCA AATAAGCATC 180

TGGATTGTCA GTAAGATAGT TTTTAAGTCT ATCTCTATCA ACTTTTCTTG GTTTTGTTCC 240

TTTTACTTGG TGGTTTAGCT CTCCTGTTTT CTCTTTTAGC TTTAACCAGC CATAAATGGT 300

ATTACGTGAG ATTTGGAAAA CGTGTGATGC TTCTGTTATA CTACCTGTTC GCTCACAATA 360

AGAGAGAACT TTTTTACGAA AATCTATTGA ATATGCCATA AGAAGATTAT ACCACATTGT 420

GTACTATTTT TGGTTCATTT TACTATATTT CTAAACACTT AGAAATAATA AAACAAATTA 480

AATATTATTT CTAAATATTT GAAAATAACA TCTATTTGTA TTATACTATC TTTGAGGTAA 540

CTATTATGAA CTATATCAAA AGACCACATT ATTTAGATTT TTTAAGAAAA CATCGTGACC 600

GACCAATCAT CAAAGTTGTG AGTGGAGTTA GACGAGCTGG TAAATCTGTG CTTTTTCAAC 660

TCTATAAAGA GGAGTTACTA GCAACTGGGG TAGACGAGGA TCAGATTATA TTCATCAATT 720

TCGAAGATTT GAGTTACTAT GATCTGCGAC ATTTTCAAAC ATTATTCGCT TATATAAAAG 780

ATCAATTAGT TAGCAAGAAA ACATACTATA TCTTTTTAGA TGAAATTCAA TATGTTGAAA 840

AATTTGAACT GGTAGCAGAT AGTCTATTCA TCTTAGCAAA TGTAGACCTC TATTTGACTG 900

GATCTAACGC CTACTTTATG AGTAGCCAAT TAGCAACAAA CTTGACTGGT CGGTATGTTG 960

AGATAGAGGT TCTTCCTTTG TCATTTGAAG AATATCTATC AGGTCAATCT CTCACAGAGA 1020

ATCTGAATAC AACAGAAATT TTTAACAATT ATCTCTTTAG TGCTTTCCCT TACTTATTGC 1080

AAACATCATC TTACGATGAA AAAATTGACT ATCTCAGAGG AATATATAAC TCCATACTGT 1140

TAAATGATAT TGTCACTAGA TTGGGAAAAC CAAATCCTAC TATTATTGAG CGCATTGTCC 1200

GAACCCTTCT CAGTAGTACA GGTAGCTTAA TATCAACAAA TAAGATTCGC AATACCCTAG 1260

TCAGCCAAAA TGTTTCAATA TCCCATAATA CTTTGGAAAA TTATTTGACA ACTTTGACAG 1320

ATAGTTTACT TTTTTATTCC GTTCCACGTT TTGATGTAAA AGGTAGAGCA TTATTGCAAC 1380

GTTTAGAAAA ATATTATCCC GTTGATTTAG GTTTACGACA TCTCTTATTA CCAGACCAGA 1440

AAGAAGACAT TAGGCATATC TTGGAAAATA TGGTATATTT GGAATTGAGA CGTAGATATT 1500

CACAAGTATA TGTTGGTAAT TTAGATAAGT ATGAGGTTGA TTTTGTTGTT GTAACTGATC 1560

TTGGCCACTA CGCTTATTAT CAGGTCAGTG AAACAACACT TGCTCCAGAA ACACTAGAAA 1620

GAGAACTTAG ACCACTAGAA GCCATTAAAG ATCAATTCCC TAAATATCTA TTAACAATGG 1680

ATACGATTCA GCCAACAGCC AATTACAATG GAATCGAGAA GAAAAGCATT ATAGATTGGT 1740

TACTAGAAAA ATAGATAAAT ATAAATCATA CAGCTAATTA GATTTGCAAC AGTCTGTTAT 1800

CAATGATTCT ACCCAAATCC TAACAAGATA TAGTGAATTT CGAATACGCT ATATAATACG 1860

GACACTTGAA AATAGAAATT GGGGATGAAA GGGGATCTAT AATTTCTGGA AGTACTATCA 1920

AAAATTAATA TCATAGTCTT ATTAGAGAAT AGCATCACCC ACTTTCTCAA ATAAGATTAA 1980

ATTGTAACTG AATTATAATG AAAAAGAGAC TGAGCAATCA GTCTTTAAAA TCAGAAAAGC 2040

GCATAGTATC AGGTATTGAA CAACCTTGAT AATATGCGTT TTATTATGGA AATATTTGCT 2100

TCATTTTCTC CTGAAATAGA GCTTTTGCTA TCCTATTTTT CTCTATTTCT AATGATTTAC 2160

TTCAACTTCT TACCTCTTGG GAAAAA 2186

(2) INFORMATION FOR SEQ ID NO: 180:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3236 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

GTCACACGTT TGACTTCACG TATTTCATAA GTATAAACTT TATTTTTATC GGTTAGATAA 60

ATCTTCATGC CATTTTTAGC ATTATCTAAA GGAGAAAATA ACATTTTATT AGCATTATCA 120

ACACCAAAGA TATGGTGACT AGCTAGACTA TAATTTCCTT CTCCCATTAC TTGCTCGCGT 180

TTCATTGTAC CAGCTCCGTA GAAGAGATTA ACATTATCAA GTCCTTTAAA AATCGGCAAA 240

TTCATTTCCA ATTCAGGAAT TGCAATTCCC CCAATAACTG GTAATTTTTG AGCATCCCAT 300

TGAGAAGTTA GAACAGCTTC CGAAGAGATA GCTTTGACAG AATCAAAGTC AAAATTGCCT 360

TCTGTATCCT GATTTTCTTC TAATTTTTCT TTTGATACCT GGCTAACTTG ATACTTATTG 420

GTATTCCAGA CTATGAAAAT ATTTCGAATT TGAGTATTAA AAATCAAAGC CAGTGACAGT 480

AATATCAGAA ATCCTGCTAG GATATTTGTC AGCAGATTTT TTCGCTTGTT TTTCTTTTTA 540

TTATTTTTTT GAGACATTAT GCTTCACCTT CTGTTTCGTT TTCTGTCCCA ACTTCTTCTT 600

TTTCTGCCAC CGCAACCGTT GTGAAAGTCA CTATCTGAGC ATCTTGATCC AGGCGCATTA 660

CTTTAACTCC CATAGTTGCA CGTCCTGTTT GTGAAATATT GGCAAGATTG GTTCGAATCA 720

TGACACCTGT ATCAGTGATA ATCATCAAAT CCTCATCCCC TTGAACAGTC ATAAGACCGG 780

CCAGCAAGCC ATTTTTTTCG GTAATTTTAG CTGTCTGCAT TCCCTTACCA CCACGACCTT 840

TTGTTGGGTA TTCAGTAGCG ACTGTACGCT TACCATATCC TTTTTCTGTG ATAATAAGAA 900

CCTCATCTTG ATCAGTAATC AAGCTGGCAC CAACAACTGT GTCTCCTTCA CGAAGGTTAA 960

CACCTTTCAC ACCAGTGGCG ATACGGCTCA TACCACGAAC GGCTGATTGA TTAAAGCGAA 1020

CTGCATAACC AAACTTGGTA CCAATGATAA TATCCATATC TCCTTCTGCC AACAAGACAT 1080

TGATTAACTC ATCTTCATCC TTTAAATTCA GCGCTTTGAG ACCATTTTGA CGAATATTGG 1140

CAAACTCCTT AACACTGGTT CTCTTCACAA TACCGTGACG GGTTGTAAAG AAGAGATAAG 1200

CATCATCACT GCGATCAGAC TCAACATTGA TAACCGTCTG AATACTTTCG TCTTCATCCA 1260

ATTTCAAGAG ATTGACTACT GGTAGCCCTT TGGCAGTCCG ACCATACTCA GGAATTTCAT 1320

AACCTTTAAG ACGATAGACA CGTCCCTTGT TTGTGAAGAA GAGCAGATGA TCATGGGTGC 1380

TAGTTGACAC TAACTCACGA ACAAAGTCAT CATCTTTCAC TCCCGTTCCT TGGACACCAC 1440

GACCCCCACG TTTTTGAGCA GTGAACTCGT CCTGATCCAA ACGCTTAATG TAGCCTCTGT 1500

TAGAAAGGGT AATCAAGACA TCCGATTCTT CAATCAAGTC CTCATCCTCG AGACTCAAGA 1560

CCTGTCCAAT CATCAACTCT GTACGGCGCT TATCAGAAAA TTTACGTTTA ACTTCATCCA 1620

ATTCGTCTTT GATAATTTGA GAAACACGTT CAGGCTTAGC AAGAATATCT GCTAAATCCG 1680

CAATCAGAGC CAAGAGGTCA TCATACTCAG ATTGAATCTT ATCGCGTTCC AAACCTGTCA 1740

AACGACGAAG ACGCATATCA AGGATAGCTT GACTTTGACG TTCAGAAAGC TTAAACTTGC 1800

TCATCAACTC AGCTTGAGcT TCCGCATcCG tTTCACTAGC ACGGATGATA CGAATCAyTC 1860

GTCGATATGG TCTAGCGCAA TCAAGAGACC TTCTAAGATA TGAGCGCGCG CTTCCGCTTT 1920

TTCCTTATCA AAACGTGTAC GACGAACAAC CACTTCTTTT TGGTGCTCGA TATAAGCATC 1980

CAAAATCTGA CGAAGAGACA AAATTTTCGG TATACCATTT TGGATAGCGA GCATATTGAA 2040

ACCAAAATTG GTTTGCATTT GGGTCATTTT GAAGAGGTTA TTGAGAATAA CATTGGCTGA 2100

GGCGTCGCGC TTGACTTCAA TAACAAATCG AACACCTTCA CGGTTTGACT CATCACGTAC 2160

TGCTGTGATA CCCTCAATGC GTTTTTCCTG AACCAAGCGA ACAATATGCT CATGCACCTT 2220

GGTTTTATTG ACCATGTAAG GAAATTCTGT TACAACGATA CGCTCACGAC CAGTCTTAGT 2280

CGTTTCAATC TCTGTACGAG AACGTAGGAC AATCGAACCT TTACCTGTTT CATAAGCCTT 2340

ATGGATACCT GATTTCCCCA TGACAAGAGC ACCAGTTGGA AAATCTGGTC CAGGCAAGAC 2400

TTCCATCAAG TCCTTGGTAG TCACTTCAGG ATTATCCATG ACCAACTTCA CTGCATCAAT 2460

gGTTTCACCC AGATTATGAG GTGGAATATT GGTTGCCATC CCAACCGCGA TACCAGTTGC 2520

TCCATTAACC AAAAGGTTTG GAAAACGCGC TGGCAAGACC AAGGGTTCCC GTTCATTGGC 2580

ATCATAGTTA TCAACGAAAT CAACTGTATT TTTGTTGATA TCACGAAGCA TTTCCAGAGC 2640

AATCTTGCTC ATACGTGCCT CGGTATAACG TTGAGCGGCA GCACTATCTC CATCCATGGA 2700

ACCAAAATTC CCATGACCAT CTACAAGCAT GTAACGGTAG CTCCACCATT GAGCCATACG 2760

GACCATGGCT TCATAAATAG AGGAATCCCC GTGTGGGTGA TATTTACCCA TGACATCCCC 2820

TGTAATACGA GCAGATTTTT TATGGGGTTT GTCTGGGGTC ACACCCAATT CATTCATTCC 2880

GTAGAGAATG CGACGGTGAA CAGGTTTTAA GCCATCTCGA ACATCAGGAA GAGCTCGCGC 2940

TACGATAACA CTCATGGCGT AGTCGATAAA ACTTGCCTTC ATCTCCTTTG TCAGATTGAC 3000

ATTCACTAAA TTTTTATCCT GCATTAATAA ATGCCTCATT TCACAATTAG TAAGTAACAA 3060

CATTATACCA TAAATTCCCA TCTATTTCAG CCTCTAAACC ACTAAAACGT TTACATCGAG 3120

AACTATAAGG CATATTCGTG ACAAAGTTTT TTAAAAGTGA TAGAATGAAG TTGTCTAGGG 3180

AAAACCCCTA ATAGAATAAG GAGATGGTTA nACAATGACT CTGACTAACA CACAAA 3236

(2) INFORMATION FOR SEQ ID NO: 181:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8651 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

AGGTCCTGAA GTATTGGAAC AGGAAGGTCA AGAGTTTTTG GAACATTTCA AAAAACTCTT 60

GGAGTCAGTT GAAGTAGTAG CCATCTCAGG TAGTCTGCCA GCTGGCCTTC CAGTTGATTA 120

CTATGCGAGC TTGGTAGAAC TTGCTAATCA AGCTGGCAAG CATGTAGTCT TGGACTGCTC 180

AGGTGCAGCA CTTCAGGCTG TTCTTGAATC ACCCCATAAA CCAACAGTCA TCAAACCAAA 240

TAATGAAGAA TTGTCTCAGC TTCTTGGAAG AGAAGTTTCT GAGGATTTGG ATGAATTAAA 300

AGAAGTACTT CAAGAACCTT TGTTTGCAGG GATTGAATGG ATTATCGTTT CACTTGGTGC 360

CAACGGTACT TTTGCCAAAC ATGGTGACAC TTTCTACAAG GTAGATATTC CTAGAATTCA 420

GGTGGTAAAT CCTGTTGGAT CTGGAGACTC TACTGTGGCA GGAATTTCTT CAGGACTTCT 480

TCACAAAGAA TCGGATGCAG AATTACTCAT CAAGGCAAAT GTCCTTGGTA TGCTCAATGC 540

TCAAGAAAAA ATGACTGGTC ATGTCAACAT GGCCAACTAT CAAGCTCTAT ATGATCAATT 600

AATAGTAAAA GAGGTATAAA ATGGCTTTAA CAGAACAAAA ACGTGTACGC TTAGAAAAAC 660

TTTCTGATGA AAATGGTATC ATCTCAGCTC TTGCATTTGA CCAACGTGGT GCTTTGAAAC 720

GCCTCATGGT TAAACACCAA ACAGAAGAAC CAACTGTGGC CCAAATGGAA GAACTTAAAG 780

TCTTGGTAGC AGATGAATTG ACTAAATATG CTTCATCTAT GCTTCTTGAC CCTGAGTATG 840

GACTTCCAGC AACTAAAGCT CTTGATGAAA AAGCTGGTCT TCTCCTTGCT TATGAAAAAA 900

CAGGTTATGA CACAACAAGC ACAAAACGCT TGCCAGACTG CTTGGATGTT TGGTCTGCAA 960

AACGTATTAA AGAAGAAGGT GCAGATGCAG TTAAATTCTT GCTTTACTAT GATGTAGATA 1020

GCTCAGACGA ACTCAATCAA GAAAAACAAG CCTACATCGA ACGCATCGGT TCTGAGTGTG 1080

TGGCTGAAGA TATCCCATTC TTCCTTGAAA TCCTTGCTTA CGATGAAAAA ATTGCGGATG 1140

CAGGTTCTGT AGAATACGCT AAAGTAAAAC CACACAAAGT TATCGGCGCT ATGAAAGTCT 1200

TTTCAGACCC ACGCTTTAAC ATTGATGTTT TGAAAGTTGA AGTTCCTGTT AACATTAAAT 1260

ATGTTGAAGc kTCGCTGAAG GTGAAGTAGT TTATACACGT GAAGAAGCAG CAGCCTTCTT 1320

CAAAGCGCAA GATGAAGCAA CGAACTTGCC ATACATCTAC TTGAGTGCTG GTGTATCAGC 1380

TAAACTCTTC CAAGATACTC TTGTATTTGC TCATGAATCA GGTGCGAACT TTAACGGAGT 1440

TCTTTGTGGC CGTGCTACAT GGGCAGGATC AGTTGAAGCT TACATCAAAG ATGGTGAAGC 1500

AGCAGCTCGC GAATGGtCGC ACAACTGGAT TTGAAAACAT TGACGAACTC AACAAAGTTC 1560

TTCAAAGAAC AGCAACTTCA TGGAAAGAAC GCGTGTAAGA AAGTCCTCCT AGTTTAGGAA 1620

CATGAATCTA AAAAAATTTA AAAAAAGTTG TATGTAAAGG CTTACAAAAT AACTTACTTG 1680

TGCTATACTT AAATCACAAG TTAATATGAA TTAGAAAGTA ACTATATGAA GTATAATAAA 1740

AATAGGATAT AGTTTATTTT ACGAGCTAGG AAGGAAAAAT ACGGAAACAA TATTGCCAGA 1800

ATAAACTATA TTTAGATGCA CATTTCATTC ATTGTTTTAT AAAAGGAGAA GATAAACGGC 1860

TACTAAAAAG AGTTTTAAAG CGTTAGTTGT AGGACTAGGT ATTGTTTCAA TATTCTTATC 1920

AGCCTTACCT ATGGTTAGTG GTTCTGTATT TGCAGATAGT GCCCTAACTA CAGTAGATAA 1980

AGCAAATGAT ATTGTTTTGA ATGTTGATGG GAATAAATTT TATAATGTTT CGGTTTCAGA 2040

AGATATTGTA AATGCTGGTC AAATTTTGGA AGATTATTTT TATGTAGATA AATTTGGAAA 2100

TATAAATTTA AAAGGCACTC CTGAAGAGTT AGCAAAAAAT ATTGGTATTT CTGTACAAGA 2160

AGCAAGTTTG ATGTATGGAG CTGTAAAAGA GTTACCCAAC GTTTACGAAA GAGGTCCTGT 2220

AGGTTTTCGT TTCAATCTTG GTCCTCAAGT GAGGGGGATG GGTGGCTGGG CTGCTGGAGC 2280

TTTCGCTACT GGATATGCTG GATGGCATTT GAAACAATTT GCGGTTAATC CTGTTACATC 2340

TGGATTTGTT GCTGTAATAA GTGGTGCGAT TGGCTGGGCT GTAAAAACTG CTGTAGAAAA 2400

TTATTGGACA GTTGCTGTAG CTACAGTAGA AGTGCCGTTT GTGAACCTTG TTTACACCAT 2460

AGATTTACCT TAGAGGTTAT TTCTTTATGA ATCATTCTTT TAAAAAAATA ACTGTATTTT 2520

GTTTTATAGT TTCTTGTGTT CTTTGTTTAT TAGACTTAAT GAATTTTAAA AATGTAGCTA 2580

CTTTTTTATT TTTCTGTCTT CCTGTTTTTG TTTTGATTTA CAAAAATAAA TAAAAACAGA 2640

GCCTCTGTTT GATGAATTTT AGAACATAGT TAAGTTTTAA AAAAAGTTGT ATGTAAAGGT 2700

TTACAAAATA ACTTACTTGT GCTATACTTA AATCACAAGT TAATACAAGG TGAGTGTTAC 2760

TAAGTAATAT TAGGCATGAT CACAGGTGAA TTAGAAATCA GCTGATTTTC TAGTTCATTT 2820

GTGGTCATTT TTTGTACTTA TATACCTTTA AGATATAAAA GGAGGTTGAC ATGTATCGAA 2880

TTCTAAATCC AATGAATCAC AATGTCTCGC TTGTCAGAAA TGATAAGGGA GAAGAGGTGA 2940

TTGTAATTGG TAAGGGAATT GCATTCGGAA AGAAGAAGGG GGATTTGATT GCTGAAAATC 3000

AGGTTGAGAA AATCTTTCGG ATGAAGACCG AAGAGTCCAG AGAAAACTTT ATGGCTCTTC 3060

TCAAAGATGT TCCGCTTGAT TTTATCACAG TGACCTATGA AATCATTGAT AAGCTATCAA 3120

AGAAATATCA TTATCCGATT CAAGAGTATC TCTATGTAAC CTTGACAGAT CATATTTACT 3180

GTTCTTATCA AGCTCTAACT CAAGGAAGGT ACAAGGATAG TAATCTGCCA GATATTTCCG 3240

CTAAGTATCC TGTCGCTTTT CAAATCGCAA ATGAAGCTTT TGAAATTTAC CGTCAGAAGC 3300

TAGCAGATCA TTTTCCTGAG GACGAAATTA TTCGGATTGC TTATCATTTC ATTAATGCTG 3360

AAGGTGAAAA TGAAGTGGAA CTTGTGGAGT CGATTGATAA GAGGAAAGAA ATTCTCAGGA 3420

ATGTTGAAGA AGTTTTAACG GACTATGCAA TTCAACGAAC TAAAAAGAAT AACCATTTCT 3480

ATGATCGCTT TATGATCCAT TTGAATTATT TCTTGGATTA TTTAGACAGA TCTAGAGATG 3540

ATAACCAATC ACTTCTGGAT ATGGAAGATC ATATTAAACA ATCCTATCCA AAAGCCTTCG 3600

AGATTGGTTC CAAGATCTAT GATGTGATTA CGCAACATAC GGGTCTTGAT TTGTATAAAA 3660

GTGAACGAGT TTATCTAGTT CTACATATCC AACGTTTATT GTCATAAAAA TTTATTTAAA 3720

ACTATATAAG GAGAATTCTA TCATGAATAG AGAAGAAGTA ACATTGTTAG GTTTTGAAAT 3780

CGTAGCCTAT GCTGGCGATG CTCGTTCAAA ACTATTGGAA GCCTTGAAGG CTGCTGAAGC 3840

TGGTGATTTT GAAAAAGCGG ACGCTCTGGT AGAGGAAGCT GGTAGCTGTA TTGCAGAGGC 3900

TCACCACGCG CAAACAAGTC TATTGACTAA GGAAGCTTCA GGTGAGGACT TGGCTTATAG 3960

TGTAACCATG ATGCATGGCC AAGACCACTT AATGACAACT ATCTTGTTAA AAGATTTGAT 4020

GCATCATTTA ATTGAACTCT ACAAGAGAGG AGTTCAATAA TGAATAAACT AATTGCATTT 4080

ATCGAGAAAG GAAAGCCTTT CTTTGAAAAA CTATCTCGTA ATATCTATCT TCGTGCTATT 4140

CGTGATGGTT TCATTGCAGG TATGCCTGTT ATTCTCTTCT CAAGTATCTT TATCTTGATT 4200

GCCTTTGTAC CAAACTCATG GGGCTTTAAA TGGTCTGATG AAGTTGTAGC CTTTCTGATG 4260

AAACCTTATA GCTATTCTAT GGGTATTCTG GCTCTCTTGG TAGCTGGTAC AACAGCTAAG 4320

TCATTGACTG ACTCAGTAAA CCGGAGCATG GAAAAAACCA ATCAAATCAA GTATATGTCA 4380

ACATTGTTGG CAGCAATTGT TGGTTTGTTG ATGTTGGCAG CTGATCCTAT CGAAAGTGGT 4440

CTAGCTACTG GATTCTTGGG GACAAAAGGT TTGCTTTCAG CCTTCCTTGC TGCCTTTGTT 4500

ACTGTAGCCA TCTATAAGGT TTGTGTTAAG AACAACGTCA CTATTCGTAT GCCTGACGAA 4560

GTTCCACCAA ATATCTCACA AGTCTTTAAA GATGTGATTC CATTCACTCT ATCTGTTGTT 4620

TCTCTTTATG CTCTTGACTT ATTAGCACGT TATTTTGTTG GTTCTAGTGT GGCAGAATCA 4680

ATCGGTAAAT TCTTCGCACC ACTCTTCTCA GCAGCAGACG GATACCTTGG TATTACCATT 4740

ATCTTTGGTG CCTTTGCCTT CTTCTGGTTT GTTGGGATTC ATGGTCCATC TATCGTTGAA 4800

CCAGCTATCG CAGCTATTAC CTATGCCAAT GCCGAAGTTA ACTTGAACCT TCTCCAACAA 4860

GGGATGCATG CAGACAAGAT TCTTACTTCT GGTACACAAA TGTTTATCGT TACCATGGGT 4920

GGTACAGGTG CGACATTGGT CGTTCCATTT ATGTTCATGT GGTTGACAAA ATCGAAACGT 4980

AACCGTGCAA TCGGACGTGC TTCAGTAGTT CCTACCTTCT TCGGTGTAAA TGAACCAATC 5040

TTGTTTGGTG CACCTCTTGT TTTGAATCCA ATCTTCTTCA TTCCATTTAT CTTTGCTCCA 5100

ATTGCAAACG TATGGATTTT CAAATTCTTT ATTGAAACTC TTGGAATGAA CTCATTCACT 5160

GCTAATCTAC CATGGACAAC TCCAGCTCCA CTAGGTCTAG TTCTTGGAAC TAACTTCCAA 5220

GTGCTATCAT TCATTCTTGC TGCCCTTCTA ATCGTGGTTG ACGTTGTCAT TTACTATCCA 5280

TTCCTTAAGG TCTATGATGA ACAAATTCTT GAAGAAGAAC GTTCAGGTAA GTCTAATGAT 5340

GAATTGAAAG AAAAAGTTGC TGCAAACTTC AACACTGCAA AAGCGGATGC TATTCTTGAA 5400

AAAGCGGGTG TCGATGCAGC ACAAAATACC ATCACTGAAG AAACAAATGT CCTCGTTCTC 5460

TGTGCAGGTG GAGGAACAAG TGGTCTCCTT GCAAATGCTT TGAATAAGGC AGCAGCAGAA 5520

TACAATGTCC CTGTGAAAGC AGCAGCAGGC GGCTATGGTG CTCACCGTGA AATGTTACCA 5580

GAGTTTGATC TTGTTATCCT TGCCCCTCAA GTTGCTTCAA ACTTTGAAGA TATGAAAGCA 5640

GAAACAGATA AGCTCGGTAT TAAACTAGCG AAAACAGAAG GCGCTCAATA CATCAAATTA 5700

ACTCGTGATG GAAAAGGTGC TCTTGCATTC GTACAAGCGC AATTCGATTA AGGCTAGAGA 5760

CTCTGAAATA GTCTCCCATC GTTACGGAAA TCGCTATGGC GAATTTCCTA TTATTAATTC 5820

GTCGGTAAAA AGATATCGTT TTTACCTCCT CATGTCACAA TTCGGTGACT TGGTACAAGA 5880

AGTGAGATGG AGAAGGATGG CTCACTGACT CCTCTCCTCT CACTTTTACT TTATTTAAAT 5940

CAAGAAATAG GTGAAAAAAA TGACAAAAAC ACTTCCAAAA GACTTTATTT TTGGTGGCGC 6000

AACAGCTGCT TATCAAGCAG AAGGTGCTAC ACATACTGAT GGAAAAGGAC CAGTTGCTTG 6060

GGATAAATAT CTTGAGGATA ACTACTGGTA CACTGCCGAA CCAGCTAGTG ATTTTTACAA 6120

TCGATATCCA GTTGACCTCA AGCTAGCAGA AGAGTATGGT GTCAATGGTA TTCGAATTTC 6180

TATTGCTTGG TCACGTATTT TCCCGACTGG TTACGGCCAA GTAAATGCTA AAGGTGTTGA 6240

GTTTTATCAT AATTTATTTG CAGAGTGTCA CAAACGTCAT GTTGAGCCTT TTGTAACTCT 6300

TCATCACTTT GACACGCCAG AAGCTCTCCA CTCAAATGGA GACTTCTTAA ACCGTGAAAA 6360

TATCGAACAT TTTGTAGACT ACGCTGCCTT CTGTTTTGAA GAATTTCCAG AAGTAAACTA 6420

TTGGACAACC TTTAATGAAA TTGGACCAAT CGGTGATGGT CAATATTTGG TTGGGAAATT 6480

CCCTCCAGGT ATCCAGTACG ACCTTGCCAA AGTCTTTCAA TCACACCACA ATATGATGGT 6540

GTCTCATGCA CGCGCGGTAA AATTGTACAA AGAGAAAGGC TATAAAGGGG AAATTGGTGT 6600

TGTTCACGCC CTGCCAACTA AATATCCTCT AGATCCTGAA AATCCAGCAG ATGTTCGTGC 6660

AGCTGAGTTG GAAGATATCA TCCACAATAA ATTCATCTTA GACGCAACTT ATCTAGGTCG 6720

CTATTCAGCT GAAACCATGG AAGGTGTCAA CCATATCTTA TTAGTCAATG GTGGTAGTTT 6780

GGATCTTCGT GAAGAAGATT TTACAGCATT AGAAGCTGCA AAAGACTTGA ATGATTTCCT 6840

AGGAATCAAC TACTATATGA GTGACTGGAT GGAAGCCTTT GATGGAGAAA CTGAAATTAT 6900

CCATAATGGT AAAGGTGAAA AAGGAAGCTC TAAGTATCAA ATCAAAGGTG TTGGTCGTCG 6960

TGTAGCTCCT GACTATGTAC CACGCACGGA TTGGGATTGG ATTATCTACC CTCAAGGTTT 7020

GTATGACCAA ATCATGCGTG TGAAGAAAGA TTATCCTAAC TACAAGAAGA TTTACATCAC 7080

TGAAAATGGT CTCGGCTATA AAGATGAGTT CGTTGATAAC ACTGTTTACG ATGATGGTCG 7140

TATTGATTAC GTGAAGCAAC ACTTGGAGGT TTTATCTGAT GCGATTGCAG ATGGAGCTAA 7200

TGTAAAAGGT TACTTCATTT GGTCATTAAT GGATGTCTTC TCATGGTCAA ACGGTTATGA 7260

GAAACGTTAT GGTCTCTTCT ACGTAGATTT TGAAACTCAA GAACGTTATC CTAAGAAATC 7320

AGCTCACTGG TACAAGAAAG TAGCGGAAAC TCAGATTATA GACTAGTAGA ATTAGTCATT 7380

AGATATAGAA TTTTAGTGAG TCAAAAAGAT GTTCAAAGAT TTTATCCAAT CTATTTATGA 7440

AAAAAAGTTT ATATTATAAA TTTCGAAAAA TGCTCTCAAA TACCGTGTTT GACGAGTGAA 7500

GAATTGAAAA GTCTTGGAAA ATGGTATGTC TCGACTGGTA AAGAATGGAT TTGTCATTCA 7560

GATGATGAGC TGGAAGAATT TAAAAATCTA TTTTTAAATT TTATCAATCC TGAAGAATGG 7620

GATACTATCT CCTTTGATTC AGATTTTATG CCGTTTCAAC AATCGTAACC AATTTCTCAA 7680

AAAAGTTAAA TCTTATATTT AGTACTCTGT AAAACTCTTA TCTAATCACG TTGCTTATAC 7740

TCAATGAAAA TCAAAGAGCA ACTTTAAACT AGGAAGCGAG TCGCAGATTT CTCAATGCAT 7800

AGCTTTGAGG AATTGGGCAA AAAGTCTTTG ATATAGAAAA ACGCATAGTA TCAGGTGTTT 7860

CAACACCTGA TACTATGCGT TTTATTGTGG GAAGATTTAC TTTTTTTCTT CTGAAATTGA 7920

GTTGTTACCC AGGCTCTTTC AGTTTATTAA GGCTTGATGA CTTTAATGTG TTTAGATAGC 7980

TTAAAAAGGA TTGAATCACT TAGTTTAGAA TCTGAAACAA TAGTATCAAG ATTTGATACA 8040

TTATAAAAAG TATAAAAATC AAACTTATTG AACTTGCTAT GATCTGCGAG TAAATATTTT 8100

TTATTAGAAT TATTTAAAGC GATGCGTTGA GCCTCTCCCT CTTCCTCGCT AAAAGTAGCT 8160

AGAGCTCCGT TTTGAATACC ATTACAGCTA ACGAAAGCTT TAGAAAATTG GAGATTAGAG 8220

AGATTTTGTA GGGTCAATGT ACCAACAAAA GCACCTGTAA TATCGCGATA ATTTCCACCT 8280

ATTAAAATCA AATCTGTTAA TTTTCGTTCG CTTAAAATCA GAAAAACAGG TAGACTGTTG 8340

GTTACGACGC GGATATTGTC AATAGGCAAC TCACGCGCAA AAAACTCTAA TGTTGTTCCT 8400

GGTCCAATGA AAATAGTTTC TCTTTCTTCT ACTAGACTGC CTGCAAAATG GGCTATTTCT 8460

TGTTTTTCTG CCGTTTGGAG GGCTTGTTTT TCAATATTTG ATCGCTCATT AGTCAAAAGG 8520

GAGTTGGTTC GAAGTTTTTC AGCTCCACCA TGCACACGAA TCAGCAAATC TTTATCAGCT 8580

AATTCCTGTA AATAGCGCCT TGCAGTCATA TCTGAAACGG CTATTTCGTC CATAATCTGT 8640

TTAACTGTTA T 8651

(2) INFORMATION FOR SEQ ID NO: 182:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3786 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

AATCTCCAAT CAGTGCCACT TCAGCTACAA AGAAGAGGAG GATAATAACT CCGTTCACAA 60

GGACAGACAA GAATAATTGA TAGAAGGAGT CGGTTTCACT TGCTTGACTT GGTCTTGTAA 120

TGATwTGGAG ACTGGCAAGC AGAATGATTC CAATGCTAAT CACACACAAG AGGGCTGTAA 180

ATCGTAGGCT ATCAAAGAAA GCAAAGAAAC TAGCAATAGC AGTGAGGAmG ATTGGAATTG 240

CCAAGAGTTG ACTATATTGT TGGAGAACCT TGTCTAGCGT CCAGTCCTTT TCCTGGTGGA 300

TAAATCGTCT CACAACGAAA CTACCCAAGA GGAATGAAAA GAAGAAGAGT GTTGTCGCTA 360

CTAGGATAGA GATGATAGAA AAAAGAGTTA AAGGAGCTAG CTGCTCAGGG AAGCGACTGT 420

TAATGCTTGC TATATGTCCA TAGTAAGCAT GTTTGATGTG ATAGATACTA AAGAAAAAGG 480

AAGATGCAGA AAACAGAATG AGCAAGAGAA AGGCTGTGTA ACTGTGTGTG ATACTTGTTT 540

CCAACTTACT TGTAGGAGAT TTGATCGCTT CCACTAGCCA AGACCAAAAA TCAAGCACTT 600

GCTCTTTCCA TTTATCCCTA GATTTTGGAG CTTGGTCGGG GATATAAGGA CTTTCTAAAG 660

ATTTACTGAT AAGAAGTGGC TCTTTCGTGG TTGCTTTTTG CTGAGGAAGA GCTTCTTGGC 720

TCTCTTCAGC TATAGTGACT TTTTCTGTTT CTTTAGAAAG GTCTGGCTCT TCTTCAGTAG 780

AATTAGATGC CTTCTTTTCT TCTATTTCTG TTCTCGCTTC ACTGTCTTCA GGAGCTTCAA 840

TTTTCTCTTC TTGCTGGCTT TCCAATTCGA CTTCAGCTTG AGGGACTTCC TCCTCTAACT 900

GAGTATTTTT TTCAATTGGT GTATCGAGAT CGGCTATCGT TTCTTCAGCC TTGTCTGCAA 960

CCTCTTGAGC TTGCTCTTCA GGCTTGTTCT TGCTTGTTGT TTTTACAAAA TCATTACTTT 1020

CAAACCATTC TTGTTTCATG GTAGAACCTC CTTTTTAGTT AGATAAATAT GTTTCCATAG 1080

TAGCAAATGT AAGCGTTTTT GTCAACGTCT GCTTGGTGTG GATATTAGAT CAATATTATC 1140

ATCAGATCTC GCAATGAGTT GATCCTTGAC ATCGGTTTTT TCAGTTTTGT AAGGGTTGCT 1200

TAATTCCGTA CCTCTTGATT CAGGCTTTTC TCTTGTGAAT TGGAAGATAG AACCATAGTT 1260

GCTTGAGATG TCCCAGTTAA TTCGTTGGCT TTCTTTCTGG TCTAGGATGA TTCTGAGATA 1320

ATCTTTGGCA GTCAGTTCAA CCTTGCCATG GACTTGGATA TTTTCAGCGT GGAAGTGATT 1380

CTCTGTTGAC TCTAGCTGAC TATCTGTAAG AACTGTATCA AAGATATTAA CGATATTGGG 1440

CGTTGTGAGT TTACTGTTTT TGATACGACT TCCTTCAATT CGGAGGATAT AGCTGTTTGT 1500

ATTGAGGGTC GCATTTTCAA GGCTAGCATT TATGATGGTG GTTTGTCCGC GATTGGCTGA 1560

GATGTTGATC CCTTTTAGAG TTCTCCCTTT TGGTAGTCGG AGAATAACTT CTTCAAAACG 1620

ACTAGAGTAG CTACTTGCGA TATGAAGAAT CCCACCAATT CCAGAAGAGA GAAACGGAGT 1680

TTCAGACAGT TTCTTATCAG TGAGACTCAG AGTTCTATCG TTCTGATTGG TGATAAGATC 1740

ATGGTGAGCA GAAAGAGATG GATGGTAAGA AATGTGGATT TGATCATCGA AAGAGTCTGT 1800

GATGGTGAGC GTGTGTTGGT GGAGAGTAAT TTCTAGGTTT TCGACTTCCT TGCCAAAGGT 1860

TAGCTTTTCC GTACGGCTAT CATAGACAGG TTCTTTGGAC ATGGAAAGTA GGCTCTTAAt 1920

CCCGTCAGAT TGGATACCTA CAAAAAGCAG GATAAAGCCG ATAACGGTAG TCACCACACC 1980

AAAGATGAGA AATCCTTTTG TCCATTTACG CATGCTGATT ACCTCTCTTT CCTTTTTTAA 2040

GAACAAATTG TACCAGACGA ACAATGAGTA GACCGAAGAA GCGAGTTGCA TAGGAAATGC 2100

CAAGTAAAAC TAGCGAAGAA GCACCGATAG CCAGTAAACC AGAACCAAAA ATCAAGATAA 2160

AGGCTGATTT GGCTTGGGCG AGGACAGTGA AACTTTCAAC TAAAAATAGG AATCCGCCGA 2220

TGATACCCAG TATGGAAACT GCAAAGAAAG CCAGAATGAC AGTCAAAGCG GCTACAAGAA 2280

TTGCGAACAG GGTCACGAGG ATGGCGATTC CCAGAGGAAT GCCGATAGGT GCTGCAAGGA 2340

GGGCTAACAA GGCGATATGT AAAATTTGTC GGTTATTTTT TTGAGCGGGT GCTTCATTGA 2400

TTTTTTTATC GAGAAGATTG GATAGAACTT CGTGGGCCGC TTCTTTGGGA GTTCCCAAAC 2460

TAGCGATGAG TTCTTCTTCT CCTTCGACTC CAGCATCGTC AAAGAGCTCT CTGAAATAGT 2520

CCATGGCTTC GATACGGTCA GCTTCAGGTA GTTTCTTGAG ATAGAGTTCT AGCTGAGTCA 2580

GGTATTCAGT TCTTGTCATG GCGGATACTC CCTTCTATGA TGCCATTGAT GGTGTCTGTA 2640

TAGAGTGCCC ATTCATCTTT TAGGGTCAAG AGCTGCTCTA TACCACCGTT TGTCAAGGAG 2700

TAGTATTTGC GCATGCGACC TTGGAACTCT CTAGAATAGG TTGTCAGAAA GCTATTGCCT 2760

TCCAATTTTT TGAGAATGGG ATAGAGTGTG GATTCTTTGA TATTAGCGAT CAGCTTAATG 2820

GTTTGGCTAA TCTCATAACC ATAAGAATCA CCCTGCTCCA GTACAGCCAA GATGAGAAAT 2880

TCAATCAAGG CAGAGGATGT TGGAAAGTAC ATGGGAAACC TCCTTTTCTA ATGTGTAAGA 2940

TTTTTATATA TAATTTTTCT ACACATACAT TGTACATCTA AAAGAAAGCC CTGTCAAGAG 3000

AAATGTGTAA AATTTTTATA TATAAAAAAC TTCTAGCTAA AACTAGAAGT TTAAAGGATC 3060

TTATCCGCTC TGTCCACTGT AAAGAGGGCC ACAGTCATCA GGATATCGAT GAGCAAGAGG 3120

GCAGCTACAG ATGGTACCCA AGAGTGGAAC AGGTCAAAAC TGTAACCAAA GAGGGTTGGC 3180

CCAAAGGCTG CTAGGATATA GCCTCCTGTT TGAGATAGGC CGGACAATTG GGCTGTCTTT 3240

TCAGGGGCGC TTGTCTTGAG TGAAAAGTTG ACCATGAGAT AAGGGAAGAG GGCACTGGTT 3300

GCGGTTCCGA TGAGGAGATG GATGGCAAGC CAGTAAATGA AATTATTGAT TGGGAAAAAG 3360

AGCATGGAAA TGCCGACCAC ACCAGCTAGT GAAACCAGAG TGAGCATGAG CTGACGGTTG 3420

CGAGTAGATA AACTGGTTGT CAGGCTTGGG ATGGTCATTG AAAAAGGAAT GCTAATCAGA 3480

GATAAGATAG AAGTCAGCAA GCCAGCTTCG TGACTGGATA GACCTGCATG GATAGACATG 3540

GTAGGTAACC AGGTCATGAC GGTGTAAAAG ATCAAGGATT GAAAACCTGA AAAGATAATA 3600

ATTGCCCAAA CCTGTTTATT ACGCATGACC TTTATTTGAC TTTTTTGTTT GGTTTGTGGA 3660

GCTAGTCTAT GATTATAGCG GTGATTTGGG AGCCAGACCA AAAAAGTTGC TAGACAGAGT 3720

AACGTGAGGA GAAGGATAAG TCCTTTCCAA GAACTGGCTT GTGTAATGGG CACAGCTAGA 3780

TAGGAA 3786

(2) INFORMATION FOR SEQ ID NO: 183:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3054 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

TCAGCTAAAA AACATTGCTA AATTGATTGA AGCTGGTGCT ACACATTCCG ATTCAACTTC 60

TCACACGGCG ACCACCAAGA ACAAGGTGAG CGTATGGCAA CTGTTAAACT TGCGGAAAAA 120

ATTGCAGGTA AAAAAGTTGG TTTCCTTCTT GATACAAAAG GACCTGAAAT CCGTACAGAA 180

TTGTTCGAAG GTGAAGCTAA AGAATATTCA TACAAAACTG GTGAAAAAAT TCGTGTTGCA 240

ACTAAACAAG GAATCAAATC AACTCGTGAA GTGATTGCGT TGAACGTTGC TGGTGCTCTT 300

GATATCTATG ATGATGTTGA AGTTGGTCGT CAAGTTTTGG TTGACGATGG TAAACTTGGT 360

CTTCGTGTGG TTGCTAAAGA TGATGCAACT CGTGAATTTG AAGTTGAAGT TGAAAACGAT 420

GGTATCATCG CTAAACAAAA AGGTGTGAAC ATCCCTAACA CTAAAATTCC TTTCCCAGCT 480

CTTGCTGAAC GCGATAACGA CGATATCCGT TTCGGTCTTG AACAAGGTAT CAACTTCATC 540

GCAATTTCAT TCGTACGTAC TGCAAAAGAT GTGAACGAAG TTCGTGCAAT CTGTGAAGAA 600

ACTGGAAACG GACATGTTCA ATTGTTCGCT AAAATCGAAA ACCAACAAGG TATCGATAAC 660

TTAGATGAAA TCATCGAAGC AGCTGATGGT ATTATGATTG CTCGTGGTGA TATGGGTATC 720

GAAGTACCGT TCGAAATGGT TCCAGTTTAT CAAAAAATGA TTATCAAGAA AGTCAATGCT 780

GCAGGTAAAG TTGTTATCAC TGCAACAAAC ATGCTTGAAA CAATGACTGA AAAACCACGT 840

GCAACTCGTT CAGAAGTATC AGATGTATTC AACGCTGTTA TCGACGGAAC TGACGCTACA 900

ATGTTGTCAG GCGAGTCTGC AAACGGTAAA TACCCACTCG AGTCAGTAAC TACAATGGCT 960

ACAATCGACA AGAACGCTCA AGCTCTTCTT AATGAATACG GACGTCTTGA TTCAGATTCA 1020

TTTGAGCGTA ACTCTAAGAC AGAAGTAATG GCTTCTGCTG TTAAAGATGC TACTAGCTCA 1080

ATGGATATCA AATTGGTTGT AACTCTTACT AAGACAGGTC ATACTGCACG TTTGATTTCT 1140

AAATACCGTC CAAATGCTGA CATCTTAGCA TTGACATTTG ACGAATTGAC AGAACGTGGC 1200

TTGATGTTGA ACTGGGGTGT TATCCCAATG TTGACAGATG CTCCATCTTC AACTGACGAT 1260

ATGTTCGAAA TCGCTGAACG TAAAGCGGTA GAAGCAGGTC TCGTTGAGTC AGGCGATGAT 1320

ATCGTTATCG TTGCTGGTGT GCCAGTAGGA GAAGCTGTTC GCACAAACAC AATGCGTATC 1380

CGCACAGTAC GTTAAGAAAA ATATAAAAAC CTATCATATC CAGCTTTAGA GCTTGTGTGA 1440

TAGGCTTTTT GTATAGAGGG TAAGAAATAG GCAAAACTTT CATAATGGAT TGATACTCTT 1500

CGAAAATCTC TTCAAACCAC GTCAGCGTCG CCTTACCGTA TATATGTTAC TgACTTCGTC 1560

AGTTCTATCT ACAACCTCAA AGCAGTGCTT TGAGCAACtG CGGCTAGCTT CCTAGTTTGC 1620

TCTTTGATTT TCATTGAGTA TGAAATAAGA TATGCACAAA TTGATTAGAA AGTCAAATGA 1680

ATTTCTACAA ATGTTTTAGC AATCGTAATG TACTTGTCTA GATTCGATCT GATATATTTT 1740

CGATTTAATG ATATGGTATT TAAAACCTCC AAAGTAGCTT ACTCCATTCT TTTACTTACG 1800

TGAGTGTAGA TGTTATTTAC TGTTTTAGCG TTTTTGTGTT CCACTCTAAC CATTATAGCA 1860

TTCTTCTCAG CTAGTGTACT AAGGAGTGTG TGCCTGAAAA TATGGGAACT AAGGGGCTGG 1920

TTTATCGGTT TCTCTAGTTT AGTATTTGCC TTTTGCAAAG TGATCTTAAA TGCCTTTCTC 1980

TAAATTTACA TATCACTATT GTTTAACAAA ATCTAATCTA TTTTAGGTCA CTTATTCTTT 2040

TTTTGAAATG TAGAATGAAC TTTTTCAAAG TTTTTCGAAT CTTTTAAAAT CTGTTTGCTT 2100

TATATCGCCA TTCTCCCCCC TTTTTTAATT CTCCCTATAT AGCCTGACAG CTTTCCCGAT 2160

GGTACGAATA TGGTTGCTTT CGTCTAGGTG GATGTCGGGG TATTCGGGAT TGAGTTTTTT 2220

TGAGGCAGCC TTGGCGGAGT TTCTTGACAT AGTTAGTGCC GTCTACTTGG AAGATGCCGA 2280

TGGTATTATA GTCAATCTGT GGGGTATTCT TGATAAATAG GTAGTCGCTG TTTCTTATCT 2340

TTGGCTCCAT GGACTTGCTG ACGACATAAG CGATTGGGTC GTAGTCGTCT GGGATAATGG 2400

AAACTCCATA TCTAAATCGT TGTCCTGCAT CGAGCGGCTA CCTGCAGAGA TAAACTACCT 2460

AACACGAGAG TAAGTAGTCT GTCTGTAGTC GTCCAGTCTG ATGATTTTTA CGATACTTCG 2520

TTTTTCTGAT CATACAGTTG CCTCTCGGCA TAGGTCAGAA CTTTACCTTG TCTGGGTGGT 2580

TCCCGTTGGT CGTAGATAGA TTGGATATCG CTAGGAGAAT CCTTTTGAAC TGGAGGAAAG 2640

AGGGCATCGA TCAAGCTACT GAATACTTTA ACTAAGTCAA ATATAGTATT TTTCTTAGTA 2700

GACCTAACCC TTTTTTCATA ATTTCTAATG GTGTTTTTAC TTATACCTAT CTTAGTACCC 2760

AATTCTTATT GAGTCCAACC ATTACTAGTC TATATTGTTT TATAGTTGAT TGAGTTTGGA 2820

ATAGTACGCT GTAGCTGCTA AAACATTTCT AGAAATTAAT TTGACTTTCC TAATAGAGTT 2880

GTTCATATCT TATTTCAATC TATTATGTTT TTCACCTCTA ACAATCGCAA TCTCTTCTTT 2940

ATCCATGAAT GAAATCGCTT TCTATTTTTG TAAGTAAAGC ATAACACGAA ATCCACGAAA 3000

ATGAAAACCT TTGTTGTGTT TTCGTAAAAA ATTTGTTGAC AGAGCACGAA ACGC 3054

(2) INFORMATION FOR SEQ ID NO: 184:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1590 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

TGTGATTTTC yGAAAATTTG GTAAAATATA TCTTAATCAT TTTCAGGAGG ACAAAAATTT 60

GACAAGATAT CAGAATTTAG TAAATGGAAA ATGGAAATCA TCTGAACAAG AAATTACGAT 120

TTATTCACCA ATCAATCAAG AAGAATTGGG TACAGTTCCA GCCATGACTC AGACTGAAGC 180

TGATGAGGCT ATGCAAGCTG CGCGTGCAGC CCTGCCAGCA TGGCGAGCTT TATCAGCAGT 240

TGAACGTGCG GCTTATTTGC ATAAAACAGC AGCTATTTTA GAACGCGATA AGGAAGAAAT 300

TGGTACTATC CTTGCCAAAG AAGTAGCAAA AGGGATTAAA GCAGCAATTG GAGAAGTAGT 360

GCGTACAGCA GACTTGATTC GTTATGCTGC TGAGGAAGGT CTCCGTATCA CTGGACAAGC 420

AATGGAAGGT GGTGGTTTTG AGGCAACAAG TAAAAACAAA CTGGCTGTTG TCCGTCGTGA 480

ACCAGTTGGT ATCGTGCTAG CGATTGCTCC CTTTAATTAT CCAGTTAATT TATCTGCTTC 540

TAAAATTGCA CCTGCCTTGA TTGCAGGGAA TGTGGTCATG TTTAAGCCAC CAACACAAGG 600

TTCCATTTCT GGACTCTTGT TGGCTAAAGC ATTTGAAGAA GCAGGGATTC CGGCAGGTGT 660

TTTCAACACC ATTACAGGTC GTGGTTCAGA AATTGGGGAT TATATCATTG AGCACAAAGA 720

AGTCAACTTC ATCAACTTTA CAGGTTCAAC TCCTATTGGA GAACGTATTG GTCGTTTAGC 780

TGGTATGCGT CCTATCATGT TGGAACTTGG TGGGAAAGAT GCAGCTCTTG TACTAGAAGA 840

TGCAGATTTG GAACATGCTG CCAAGCAAAT TGTTGCGGGA GCCTTTAGCT ACTCAGGACA 900

ACGTTGCACG GCCATTAAAC GTGTCATTGT TCTCGAAAGT GTAGCAGATA AATTAGCTAC 960

TTTGCTTCAG GAAGAAGTTT CTAAATTAAC AGTTGGTGAT CCATTTGACA ATGCTGATAT 1020

TACACCTGTT ATTGACAATG CTTCAGCCGA CTTCATTTGG GGCTTGATTG AGGATGCACA 1080

AGAAAAAGAA GCTCAGGCTC TTACACCAAT CAAACGTGAG GGCAATCTTC TCTGGCCAGT 1140

GCTTTTTGAC CAAGTTACAA AAGATATGAA AGTGGCATGG GAAGAGCCAT TTGGTCCTGT 1200

TTTACCAATC ATTCGTGTGG CTAGTGTAGA GGAAGCTATT GCCTTTGCCA ACGAATCTGA 1260

ATTCGGCCTT CAATCATCAG TCTTTACAAA TGATTTCAAA AAAGCCTTTG AAATTGCTGA 1320

AAAACTTGAA GTAGGTACAG TCCACATTAA TAATAAAACC CAGCGTGGTC CAGATAATTT 1380

CCCATTCCTT GGTGTCAAAG GTTCTGGAGC TGGAGTGCAA GGAATTAAAT ATAGCATTGA 1440

AGCGATGACA AATGTCAAAT CCATTGTTTT TGATGTGAAA TAACGTGTAA AACCAGGAAA 1500

TTGTTTTCCT GGTTTTATTT TTTTGCTATA AAATAATAAT AATTATAGAA AAAATACGAA 1560

CTTTTTGGTA TTATAATAGA TTGAAACCGG 1590

(2) INFORMATION FOR SEQ ID NO: 185:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4848 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

CCTGCAGTTG TCAGACCTGT AATTTTCTTT TTATCTGTAA TAAGAATCGT TCCAGCGCCT 60

AGAAAACCCA CACCTGATAT AACTTGAGCT CCTAATCGTG TAGGATCTCC TGTCCCAAAT 120

TTATAAGATA CGTATTCATT CGTCATCATA ATCAAACATG CAGCTAGACA AACAATACTA 180

TAAGTTCGGA TGCCTGCAGG CTGGGATTTG CTCCCTCTCT CTAAACCAAT TATACTACCA 240

ATGACTACTG ATAAAACAAT CCTGACAACT ATTTCAATAT TTGATAACCC AAGACTAGTG 300

GCTGTCATGA TTATTTCCTT ACTTTACGCC CCGGTCTTTG TGTGAAGTAT AATACCGTTC 360

CAGAAATAAT CATCAGAACA ATTGTATAAA CAAATACCAG AGCTTGTGCA TTAGATGTTG 420

CTGTTTCATC ACCTGCAGAT CGAATCGTAA TACCTAATGG TTGAGCTAGG GGATGGTAAA 480

GGAATACAGA TAAGTCGAAG TCAGTTAATA AAGAGTTAAA GTTTAAAGCA ATAACAGAGA 540

GAACAACCGG TAAAATAAAT GGAATGATAA CCTTCATCAT AGTATAAAAA GGTGAAGCAC 600

CCATACTTCT TGCTGCATCT TCCATCTCAT CATCAACACT AAATAAAATA GCACGTACCA 660

TTCTATAAGA AAATGGGATT TTTACAACTA TATATGCAAT AAGTAGAATT ACCAAACTAC 720

CTACCAAAAT CTGATTCAAG ACAAGAAATT GTGGCTGATT AAAAGTAAAT AATAAACTTA 780

CTGCTAAAAG TGTACTTGGT AGTAACCAAG GAAGTAGAGC ACCATATTCA AATAAGAAAT 840

CAAAACGAGA TTTATGTTTT CTGACAACAC GAGCAAATAC AACTGCGAGA ATTGTTGCTG 900

TTGTCGCAGC AATAATAGAA TAAATAAAGC TGACCAAGAA TGGAGAGAAT GCCGCACTAT 960

TACTAAAGAA TAAGCGATAA TTTTCTAAAG TAAAGTTTGA TAATGTTAAG TTACCTGTTT 1020

GAATTGCAAC TGGATCTGTA AATGAGTATA ATACTATAAA AATTAGTGGA AGCATGAAAA 1080

CTGTGAACAA TCCATATGCT ACAATGTGAG CAATGATATT CCAAGGCTTA GACGCAATTT 1140

TTTGTTTTTT AAGAGGCGCT TTAGTCTTAG AGATAGAAAT ATAATTTCCA CCTTTTTCTA 1200

TCTTATTCAT GATAGTAAGC AAAATTGTAG TTGCAATACC TAAAATAATT GCAAGTAGGG 1260

CAGCTAAATC ACGAGAATTC CCCATCCCTG CAAATGTAAT AATCATTGGA TTTATAGTTT 1320

GAAATTCTTT ACCACCAACA ATCATGGGTG CTGCTACTGC AGATAAACCA CTAAGAAAAA 1380

CCATAATAGT AAGTGCAAAT AGAGTTGGAA TTAAGGTTGG TAACACTACT TTTCGGAAAA 1440

CAGTAAATGG TTTTGCTCCC ATATTTCGAG CAGCCTCAAT AGTGTGATAG TCAACGCTTC 1500

GAATTGTATT TGTTAAAAAC AATGTATGAT TAGCAGTTCC TGAAAATGTC ATAATGAATA 1560

AGACTGCACC ATACCCAATA AACCAGTTAG GGTCTAAAGA AGGGATAACA TTTTGTAAAA 1620

ATTTTGTAAT CAATCCATAA GGACCATAGA CAAATTTATA TCCAGTCGCT AAAACCACTC 1680

CTCCATAAAT TAAAGAGGTC ATATAACCTA ATTTTAAAAT TTTAGCACCT TTAATATCAA 1740

AGTACTCTGT AAATAGAACA CAAAGAATAC CTACGACATT AACTGTAATA ATGAGTGAAA 1800

ATGCTAACTT AAAACTGTTC ATAATACTCT GAAGTGCCCT CTGAGATTTT AGAACACGAT 1860

GTACAGCATC AAGGGAAAAT TCTCCTCCTT TTACAAATAC ATTCACTACT AGATCAAAGT 1920

TTGGATAAAT AATAAATGTT ACTAAGAACC AGATTAACCC TAAACGAATA AGCCAATCTT 1980

TTAAATTTAA TTTATGACGC ATACTGCACC TCCTTAAAAT TGCAGAACGT CTGATGGTGT 2040

GATAAATAAT TCCACACTTT CTCCGACAGA TCTAATAGCA GCCTGACTAT CAATACTTGT 2100

TACATTAAGA ATCTGACTTT CAGAAACTTT TATTGTATAG TGAATTGTAA CTCCAGAAAA 2160

CTCAACATCA ATAATTGTCC CTTTTAGAAT AAAATCTTGT TCAGTTTCAC GATTGAATCG 2220

AACTTTCTCT AATCGAATGT ATCCTTTTTT ATCCTCTAAG AAAACGCTTG TATTTTTCAA 2280

TAATACTTCG TGGACTGTTT CATCGGTCAA AACATTAATA TCTCCAATAA AATCACATAC 2340

AAATTCAGTT TGAGAATTAT GATAAATCTC TACTGGTGTA CCGACCTGTT CGATGTATCC 2400

ATTGTTAAAG ACTGCAATTC TATCAGATAA AGTCAAGGCT TCCTCTTGAT CATGAGTAAC 2460

ATATAAAGTA GTAATACCTA ACTCTTTTTG AAGTCTTTTC AACTCTTTTC TCAAATCTAC 2520

ACGTAATTTT GCGTCAAGGT TTGACAATGG TTCATCTAGA CAAAGAATTT TAGGTTCAAG 2580

AACCAGAGCA CGAGCCAATG CTACCCTTTG TTGTTGACCC CCAGATAATT CTGATACATT 2640

ACGCTGTAAC TGTTGATCAG AGATCTTAAT TTTTGCTGCC ACTGCTGATA CTTTAGCTTT 2700

AATAACATCT GGAGCTACCT TCTTAACTTT TAAACCAAAT GCAATATTAT CAAAAACAGT 2760

CATAGTTGGA AATAGCGCAT AAGATTGAAA TACAATACCA ATTCCACGCT TTTCAGGTTC 2820

CAAATGAGTG ACATCTGTTC CATTAACTTC AATACTTCCT GATGATGGAT CTAGAAAACC 2880

TACCAATGCT CTCAAAGTAG TTGATTTACC ACATCCTGAA GGCCCAAGAA ATGTAAAAAA 2940

TTCCCCTTCA TGTATATCTA AATTCAGATT ATCAATTGCA ACAAAATCAC CATATTTAAT 3000

TTGAATATTA TCAAATTTAA TCATCTCACT AACTCCCTCT ATTACTAAAC CAAAAGCCTC 3060

TCTTTATTTC TTCCATAAAT TTAGAAATAA TAGAGAGACT TGGACATAAA AATTAACTCT 3120

TATTTCTTAT TGTACGTATT CTAATTCAGC TTTTTCTACC CATTCATCCA AATGCTTTCC 3180

AACAGCTTCC CAGTCAATAT TTTGTGGTTT CACTTGATCA ACAAATTTCT TCGTATCTTC 3240

AGGTAGATCT TTGAGGGCAT CTTTATTTGC AGGAATAGAT CCAAAGTTCT TACTATATTC 3300

TACTTGAATT TCTGATTGAC CAAACCAATC AATAAATTCT TTAGCTAACG CTTGTTTTTT 3360

ACTAGTGCTT AAAACCATAG TTTGTTCAGT TACAAATGGT ACACCAATCT CAGGAGTCAT 3420

AACTTTGAAA ACAACATTTT GTTCTTTTTG TCCAACTAAT GCACCAGAAC CCCACATCAT 3480

TCCATATTGT ATTGGATCTT CTTTGTCTAA CATCTTAACA ATTGAACTTT CTCCCTTTTG 3540

AAGAGTGTAT GCATTTTTCA AATATTCTTT TGCTACTTCC CAACCTTTTT CGGAAACACC 3600

TAATTCACCT TTATCATCAA GGTATCGAAC TAAGATACTT GCTAGAATTG CCCGTCCTGT 3660

ACCTCCTTGA AGACCAGAAA TTGAATATTT ACCTTTATAC TTACTACCTA ATTCAGTCCA 3720

ATCTTTAGGC ATTTCTTTTA CATCAGGCGC CCCAATTAAA ACTAATGGTT GAACAATCAC 3780

AGGATTATAA TAATTATCTT TATCTGATAA AGATTGATCA ATTTTATCTA ACCATTTAGG 3840

CTTGTACTGT ACTAGTAATT TTTGATCTCT AATTTTATTT GAATCAACAG CACCAATTCC 3900

AAATACCATA TCTGCAACTG CATTATTCTT CTCAGCAATA ACACGGTCTG CTAATTGAGC 3960

GCCAGCGATA TCAACCATTT TTATATTAAA ACCAGCTTCT TTTGCTTTAG CAGTTAACCA 4020

ATCACCACGA CCATTTGAGA CTGAGTTCGA ATAGATAACT AATTCTTGAC TTTTATCAGC 4080

TTTTTCTTCA GATGAAGAAG CAGTCGTAGA ATTTGAACCT CCAGAGCAAG CAGCAAGTGT 4140

AGTAAgAGCA ACTCCCGTTG CAAGTACAGT AGACCAAACT TTCATTTTTT TCATGATAAG 4200

TTCTCCTTTT TTATTATTTT ATTTAAATTT TTCGTGATAT GGAACAAATT GTCTCATATC 4260

TTCAAATACA GTATAGTCAA TACGGTTTAC AGTAATAGTT GGAATCTTCT CTAATAAAAT 4320

TTCAGTTAAT TCTGCTCTGA CTTTAGTAAA CTCTTCTTCC TCCTCTTCGG TTAGAGGAAT 4380

CCGAAGATAC CCAATTGAAA TATGGAATTG ATATCTATCA TGATTAGGGA AACAAACACC 4440

TGCTTTTTCT GAGACATAAG TACGAATTTC TTCTAATCTC TTTGCAGAAG CTTCATCTGC 4500

AGGTTCAACT AGTATGTTTT GTTTTCCCAT TTCAGTTATA CGCATATGAA TTTCTTCATC 4560

CAACAATGGA AAAATTTCAA GTTGTTTAGC AAAGTAATCA TGTATTTCCT GTAAAGGTGT 4620

ATCTAGAGGA AGATTACTGC TCCAAAACTC gtTCACGATT TTCATGGCAC AACAATTCAA 4680

TTACAGTCAT GTGAATAGAA TTCCTTGGAG TTAAAGTAAA CTTATCGATA AATGGTAATT 4740

CTCTATAACG TGATTGAATA ATATCAACAA CTTCCATCAA ATCTTGTTTA GTATAAAGAT 4800

TTGCTACAAC TGTATTCCCA GGGAAATGAT TAAATTCCCC ATTCTCGG 4848

(2) INFORMATION FOR SEQ ID NO: 186:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3763 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

GTTATAAGCA ACACCTTCTT GCTTGCCATA AGTTGTGAAA TGGGTAGAAT CGATATCTAC 60

AATGAGTTGG TTTAGCTGGT GAAACTGTAA AAAGAATTCG ACCAATTCAA GGTTGAGGCA 120

TCGCAAACTA TGGACTGTTT CCTCGTCAGT TCTGGAAAGA AAACGGGATA AGGTTGGCTG 180

TGAAGCAAGC TGCCCTCCTT CCAATAATTT TGGAAAGTAG GCATCAGCTG ACAATTCTTT 240

ACAAGCATAG TCCGTTCCAT AACCTGTTAA CAGTTGAAAG AGGAACTGGA CAAGGATATC 300

TGAATCCGAA TAACGACAGT AGCGGCGTTG GTCATTCGTT ACTAAATACT TAGAAATCCG 360

CTCTTTTAGT TTCAACTGGG AAAAAAGTTC CTGAAAAAAG ATAAGACCAC CATACTGGGT 420

TAAATGACCT CCATCGAAAG ATAGTTGGTA AAAAGACTTG TTTTGGAAGT GATGATTTGG 480

TAAACTGTTC ATGTGAGTTT CCTTTCTTTT TGTGTTTTTT TCTACACTTA TACCATAAAG 540

GGGAAACTCT TTTTTGTCTA GTAAAAAACA CCCATTGGGT GAAAAAAGAA ACCATCCAGG 600

ATCTAAGCTA AGGCAAGGAT TCTGGATGGT TTTTAGATTT GGGGTGAATA ATTGGGGATT 660

TAGGAGAAAT GATGGTATCT TCCAAATCAA AATCAACTTC ACTCCATAGT CTCAACTGAT 720

TGATTTTCCC ATCTTGATAG GTCACATCCT TGTCAAGGAT AAACTGAGTC AACACCTCAT 780

GTTGACCTTG ACACCTGATG TCATCTACCA AGAGCCAGAC ATCCTCTACC AACATGAGGA 840

TTTTTCTCCT GTGAAGATAA GGCAAATCAG GTTCTGCTGA CCAATAAGCC CCCTCAATAT 900

AATGCACTCC CTCCCTTTCT TTATGGTGAC AAAACAGGGA GTGAGGATAG TATTCATATT 960

CCCAGGATCC CGTGATTCTT TCCGGAGCTT TCCCATCTAC AATGCAGGTC GAATGACTCC 1020

AAGCACTCTT TAAGAGATAA CGTTCATATA TCTCCCGATA AGAATAACGC CCAGCATCTA 1080

TGAAAATAGG TTGGCCTTGA TACTGTAAGC AAAAACTATT CTCGTCACTA TGACTATGGG 1140

CACTTCCTAG CGGACCATTT TTGAAAAATA GATAACGATG TTCATCCTTA ATGCAGACAT 1200

GTCCAGAGTC TTCAAAGATC ATGGACTTAG GCTGCCAAGC TCTCTTTTCA AATTCCTGCA 1260

GTCGCTTGAC CTTTTCTCGC CCCAGGAACA AGAGGCTAAG CAAATCAACT TTAACATCCA 1320

GACCGTTAAG AAGGTCTTCC TGGTTCAAAA CCACAGCAGA CAGGCTCAAA ATTTCTGTCG 1380

TTTCTGTAGA ATCGCTATCA CCAAAAGCCA AAGTCCGTCC ATCTAAGCCT GTCATCATTT 1440

GAATATAGGT CGCCATCTTT TCCAGCAACT CTTGGTAACT ATCTTGCAAG TCTGGAAGCA 1500

AGAGACACAA ATCCAGCAAG GCTTTATAAA CCTCTACATG ATAGAGAATC GACTGTTCAA 1560

ACTGGCTTCC ATCTCCTAAA ATCTGTGTCT CAATTTGCTG TTTCAACTCC TCTGAAGCAA 1620

AATGGTAAGC TTCTTCTAGA TCCATCTTAT CTGAAAAGAA ATGATAGATA GCAAGCATCG 1680

GAATTGTTTG TAAAATCCCC CAGTTACTAA GGGTGTACTT GGCGCGATAG TAGCTTTTCA 1740

TAAAGTCAAT CTGCTTTTCT AGACTGACCA AAATTTTCTC TAGTTCTTTC TCCTCTAGCA 1800

AGTCAAATTT CAAGAGGAGC AAGAGTAGTT TCAACCAAGT AAAGGAACGA ATACCCGTAT 1860

CCAAGGTTCT AGTCATCAAG GATTGAGGAG AAAATTCTCT CACCTGCTCA ATCCAATCAA 1920

ATAGAAAGAA CTTGCACTTT TGAATATAGT CCTTATCTCC TTCTACCAGA TACCCTATCA 1980

TAAACTGCAA GAGATATTCT TGTCGATTGA GCATATAAGA CCATTCTGGA TCATCTTCAA 2040

ATACTTGATC CCATACCATC GGCTGGATTT GATGGATTTT TGAACAAGGC TCCATATCCC 2100

AAGGACTATC AAACATAAAA CGATTGTCCA TCAAGCGTTC AAGGGAACTC TTGACTTTCT 2160

CATAGTCTTT TGAACAGTGC GACAAGATAT AATCACGACA TTGATTTCCA TCGACTCTTT 2220

CAAAAAATTG TCTTCTTTCT TCTTTCATTA TCTATTACCA GAAAAAGAAC TACTTAAAAA 2280

GCAGTTCTTT TGTCTTTCCC ATTACACTTT CCTTTTCTAC ATGGATGACC ACACCTTTTG 2340

CAATCTGCAA GGAGACCAAG TCATCTTGGA TAGAAATGAT TTTTCCATGA ATTCCAGACA 2400

ATAACAACAC TTCATCACCA AATGTTAAAG AAGCTAAATA CTCTTGTCGT TGCTCCATCT 2460

GTTTGCGAAG CAACTTTTGC TGACGAATAG AATGAAAGCT TGACAGTAAA AGGGGACTCA 2520

CTGCCAAGAC AATCACTATT CCATAAAACA ATGTTGTATC CATTAAGCTA TAATCTTAAG 2580

CCAGCTTCCG ATAATTCCGA TGATAACTGT TAAAATAACG AGTTTATATG TTGTCCATTT 2640

CTTTTCTTTG ATCAAGTAGT AAACTAAAAG TGTAAATAGG GCTGGTAGAA GAGCTGGAGC 2700

AACCTTATCA AGCATTCCCT GAATACTTAC GATACTTTGT TTAGCGTCTG CTTTAACTTC 2760

CCCTGCAGCA AAGGTAATCG GCACCATAAT CTTAACAGAT GTCGCTGCCA AACCAGCAAT 2820

TACGtTACAC CGATAATATT GGCAATACGA GAAATCGTTG CCATCTGTTC GCTTAGTTTA 2880

TCAATCACAG TTGTTCCTAG TTTGTATCCA TACAGACCAG TTGACAATTT AATCGCTGTT 2940

AAAATCGTAT TCATCGCAAG GAAGAACAAG ATTGGACCGA CAACCAAGCC TTCTTGAGCA 3000

AACGAAGCTG CGATGGTTGA GAACAATGGA GCTAAACAGA ATTGAGAAAG AGAATCCCCA 3060

ATACCTGCCA ATGGTCCCAT CAAGGCCATC TTGATGCTAC GTGTTTCTTT TGCCGGACGG 3120

CCATTTTCCA ACATTACAAG ATGCAAGCTG GTAATAAAAG GCAGGAAGTG TGGGTTGGTA 3180

TTATAGAATT CACAGTTTTC TTCCAAGGCT TGGTAGAAAC CTTCCTGATC CTCTCCATAG 3240

TGTTTTTTCA AAGCAGGATA CATCACATTG GCATATCCCA ACCCTTGATA GTTACTATAG 3300

TTAAATCCAT TTTGACAAAA GAATGCCCGC AAAGACGTTT TAAGATAATC ACGTTTTGTT 3360

AATTTGTTAG ATCCAGTCAT CGTGTGCTTC CTCCTCTACC ACATGATCCG CTGTTTTTGG 3420

CTTGTTATAA AATTCAATCA AAGCAAAGAT AGTACCTACA ATTGCAATAC CAATTGTTGG 3480

GATGTTTAGA TAAGCTGCAC AAACATATCC CAACAAGACA AAGGGAATCA ACTCTTTCTT 3540

AGCCATCACT GACAAGATCA TCGCAAAACC GATAGCTGGG AGCATTTTAC CAGCAACTGT 3600

CAAACCTGTA AGTAATACCG GTGGAATGTA GTCTACGAGT TTCAACAAGG TATCCATTGA 3660

AAGGGCACCA AGCAACCCAA GGTAAATCCA ATAAAGGCAA ACAACCAAAT TGTTGCATTT 3720

AGAGTGAACT TAAATTTCTT CAAATTATGG TTTTTCAAGT GCT 3763

(2) INFORMATION FOR SEQ ID NO: 187:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5053 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

CAATCTCTGA GTATGTGCGG TCAATACTAw CAAAGGGAAT yCCTGACGTC AAGTAATGTT 60

CAATTGGmCT ATAGGTAATG GCAACCACTC CATCAACTTT ATTATGACGC AACATCTCCA 120

GATAGTCTTG CTCTCTATTT GTACCATTGA TAGAACATAA GAGTAATTTG TTATTTCTCT 180

TATAGACTTC ATTTTCCACA TGCATAGCAA ATTCTGAAAA GAAGGGATGC CAGATACTTG 240

GTACAATGAT TGCAATCGTT TCTGTTCGAT TTTTTTTCAT TCCTCTAGCG TAGTAATCTG 300

GAATGTAATT CAAAGTTTTA ATCGCTTGTT CCACTTTTTT CAAAGTTACT TCTTTAATGC 360

CTTTTTCTTT ATTAATTACA CGTGAAACAG TTCCAACACT AACTCCTGCT TCTAAAGCAA 420

CATCTTTCAT GGTAATTGAT TTTCTTTGTT CTACCATATT ATCACCTCCT TTCAATATAT 480

AGTATCATGC AAATGCTTTT TAAGCAACTA TTTCTCAATC ATTTTTGGCC AGATCATTTA 540

TCCCATCATG AATAAAATCA CTCCAATTAG CTTTTGAAAA TACTTCAATT TTCATGTGTA 600

AACATCTACA TAAAACAGGA AAAGCCTTGG TTTCATGGCT TTTTTCGTAT CTTCTATAAA 660

AAAAGCAAGA GTTTTAGATG GCTATAAATC TAGATGTACA TTTTGCTTAA ATGATTGAAG 720

GTCTTTTCTT AACAAAAACA CCCCCAAAAT TAGACTTTTT CTGTCTAACT TTTGAGGTAC 780

AGTTCAAACG CGAAATAGCG TTTTTTTGTT ATTTTTGGTT ACTCATCTAA TCGAATAAAC 840

ATCATGGCAT TTAACAAGTA TATGAGTGAG ACCGTGTTTA TATTATTTGA ATAGATGAGT 900

CTCTTATTTT CAATAGGAGG AATAATAAAA TTAGAAATAA TGATATCATA AGGTGAATCT 960

TCTAAAGATT CCTTTGATAA TTCTAATTCA GTCCAAACTT CCAGTTCAAA ATTATTGCTA 1020

CAATAATAAG AAAGTGTCTC TGCAACGAAT TTTGCATGAT ACTGATCAAA ATTACTCATA 1080

ACTAAAACCT TTAGTTTAGG CTGATTTTGT AGCAAATTAA TCACCAAATG TTTGGTATGA 1140

GTGATGAAGG TATAAGATAG ATGATTTACC ATCATTGAAC TAGAACAAAC CTCAAGAGTC 1200

TCTAAATAGT GAGAAAGCTC TTTTTTTATA TCTGAAACAA ATTTTGGAAA AATATTTTGA 1260

AAGTTCCTGA TTGTATTCCC TTTTTGATCA AATAAAATAA ACTCAGTAAA CAACTCTTGA 1320

CGATACAGAT GTGCGGTATT ATGCAGATGC CAAATCAGAT TATCCTTATT CTCCATTTCA 1380

ATCTGATACT TGACTGAAAT CTGATCAATA AAATCACTCA ATAGATGGTA AGATTTTTCA 1440

ACATAACTAT CCTTTTTTAC GCATTTCATA AAGAGACTTT CATCTATGAA AAACATTTTT 1500

TGAAAGTAAG ACACAAATAA TTGGCAAACA ACTTCTTCAT CTAAAGAGAT ATTGTATTCT 1560

GATTCAAAAC TCTGAGCAAC ACCTTCTATT CCTTCTGCCT GCATTAAAAA ATCCAAACTT 1620

TGGTCGTTAA AAGAATCTTT ATCTACTTCC ATAAAATGAC CAAACTTTAT TCTATATAGG 1680

TTCGTAACTA GGAGCAACTT TAGCATTCTA TGCGTTGACA AATTCATTGG AAAGCTTGTT 1740

TCCTTATAAA CCAATTCTAA CAATTGAGAT AGTGGCTCTG ATGAAAAATT TTCAAATGGC 1800

CATTCTAGGA AATAATATTT TTCTGAAAAA TATTGTGCAA AAAAGTAACG AATGTCTCTC 1860

TCATTTCCAA TGATTTGAAC AGGGGTCAGA CTAACTTCAA ATTGAAATTG CCTTTTAATC 1920

ACTTTATTGA TTTGGCTAAT AATACGATAG AGCGAAGATG AACTGATATA AAATTCTTTA 1980

CAAATACTCT CAGCTTGACA ACCTTCATTA AAGAAGATGA ATTCTAAAAT CGAAAAATGA 2040

GTTGAATGTT TAAAGAAATG ATGGTAAACC ATTTCAATAT CACTATCATC GGTATTAATA 2100

ATGCGTATAC CATTAGTAGA AGAATGAAAA ATCAAGTCAG GAAAAGCAGA TTTAACATGG 2160

GATAGATCAT CTTTGACTGC ACGTTCTGTA CAATTTAATA ACTCTGCTAG TTCAGAACGA 2220

TGAAACCAAC GTTTATGTTC AAATAATAAT TCTAATAATT CTAATTGCCT ATGACTTTTT 2280

TTAGATAATA AATCTCTCAT GAATATCTTT CTCTCTTTAT AAATTATCGG ATTAAACCTC 2340

TTGCAATTAT ACCACAAAGA ATAGGTATAG CATGATATAA CGACTTTTCC TAAAATCTTT 2400

TATTTCGTAT AATAACACTA CGGAGACAAT ATATAAACAA TTTTCTTATT TTACCGTCTA 2460

TTGAGGGCGT GAATACAGAA TCAAATTCAA GTCTAAAGAT TATATTTTTA ATTTTAAAAA 2520

TTATATAATA GCAACAATTA AAGAATTTGA TTTTTTAAAA TTATATAATA ATAACAATCG 2580

AAATAATTGA CTTTTCTATA TTAAAGTTAT ATAATAGTAA TAATCAAAGA AATTGATTTT 2640

TTGATATTAA AATAAAAAAG GAGGGTAGGC AGTGTTGTGA TCAATTATTG CTGGAGGTCT 2700

TATTGGTCTC TTGGCAGGTA AAATCACTAA AAAAGTAGTT CTATGGGAAT CATCGCAAAT 2760

GTATTCGCTG GTTTAGTCGG GGCATATGCA GGACAATCTC TTTTAGGTAG TTGGGGTCCA 2820

GCAATCGCTG GAATGGCTTT GCTCCCATCT ATTGTAGGTG CAGCGATTGT GATTACTGTA 2880

GTGTCATTCT TTACAGGTAG AAAGTAAACT TTTCGCCAGT AAAGTTAGCA AACTATTTTT 2940

AAATCAATGA CGGGAAAAAT AGTTTAAATG TTAAATCGAA AGGATTGTAT ATGTCAAAAG 3000

CAAAGAAAAT ATGTTTCATT ATTTTCTGTA TTTTAATCTT GACAATTTTC CTTCCTGTTT 3060

TGATAGATTA TCATCAAGTT AGTGATCTAG GTATTCATCT ACTTAGCTGG AGACAGAACT 3120

CCGTAGTTGA ATTCTATCTT GCTAGATATG TCTTTTGGGG GACAGTGGTT CTATCAACTT 3180

TAGTTTTATT ATCCATTTTA GTTGTGATGT TTTATCCTAA ACGTTACTTG GAAATCCAAC 3240

TTGAAACTAA AAACGATACA TTAAAATTAA AGAATTCGGC AATCGAAGGT TTTGTTAGAA 3300

GTTTGGTGAG TGATCATAGA TTGATCAAGA ACCCAACTGT TCATGTAAAT TTACGAAAAA 3360

ATAAATGTTT CGTTCATGTA GAAGGTAAAA TTCTTCCTTC AGACAACATC GCTGACAGAT 3420

GCCAAATAAT TCAAAATGAA ATAACTAATG GATTGAAGCA GTTTTTTGGT ATTGAGCGTC 3480

AAGTAAAACT TGAAGTTGCA GTAAAAAATT ACCAACCAAA ACCTCAAAAC AAAAAGACTG 3540

1140

TTAGTCGTGT GAAGTAAGGA AGTAAAAAAT GGAATGGCTT AAACAATATC GATATCCAAT 3600

TATCGCTGGT CTCATAGGCG TATTTCTGGC TTGTTTGATT GTCTCCTTTG GCTTCTTCAA 3660

AACAATATTT GTATTGATTT TAGGAGCACT GGGAGTTGCA GCTGGATTAT ATATCGAAAA 3720

AAACTATATA GATAAATAAA AAAATAAAAA TTACTAATTT AATTAAAGGA GTTTCATATG 3780

TCAAACGAAA AAAACACAAA CACTAACGTA GAAAAGAAAG ATGCTACTGT TGTAGCTCAC 3840

GAAATCAAAG GGGAACTTAC TTACGAAGAT AAAGTTATCC AAAAAATCAT TGGTCTTTCA 3900

CTAGAAAACG TTTCAGGTCT TTTGGGAATC GATGGTGGTT TCTTCTCAAA TCTTAAAGAA 3960

AAAATCGTTA ACAGCGATGA CGTAACAAGT GGTGTTAACG TAGAAGTTGG TAAAACACAA 4020

GTTGCAGTTG ACTTAAACGT TATTGTTGAG TACCAAAAAA ATGTTCCAGC TTTATATTCA 4080

GAAATCAGAG AAATCGTATC TTCAGAAGTT GCTAAAATGA CTGACTTGGA AATTGTTGAA 4140

ATCAACGTAA ACGTTGTCGA CATCAAAACT AAAGAACAGC ATGAAGCAGA CTCAGTAAGC 4200

CTTCAAGATC GCGTATCTGA CGTTGCTGAA TCAACAGGAG AATTCACTTC AGAACAATTC 4260

GAAAAAGCTA AATCTGGTCT TGGATCTGGT TTCTCAACTG TTCAAGAAAA AGTTAGCGAA 4320

GGTGTAGAAG CTGTTAAAGG TGCAGCAAAT GGTGTAGTAT CTCACGAAAA CACTCGTGTA 4380

AACTAAGATA AAATAAATAT AACAGGAGAA ATTATCATGT CAGTAGAAGA AAAATTAAAT 4440

CAAGCTAAAG GTTCTATTAA AGAAGGTGTT GGGAAAGCCA TCGGTGATGA AAAAATGGAA 4500

AAAGAAGGTG CAGCTGAAAA AGTTGTTTCT AAAGTAAAAG AAGTTGCCGA AGACGCTAAA 4560

GACGCTGTAG AAGGTGCTGT AGAAGGTGTT AAAAACATGT TGAGTGGCGA CGATAAATAA 4620

GGTTAAAAGT TACTTTATCT TTTTAGTAAT ATTAGTCAAA AGAGTCTGAG TCAAGATGAT 4680

TCTCAGAAAA CAAAAAGCTA GAGATTCCCA ATTGCGGAAC TCTAGCTTTT TAATTTTGCC 4740

TCTTTCTCTT ATTATATTTC AGCAGGTTGT TGGCCATGAG TACGAATCCC ATGTCAATTC 4800

TCACTTGACG CTTACCTCTC AGATGACATC TCTTATAACC CAAACAAACC TTTATCTGCC 4860

CAAAGACAGA TTTCATATCA ATCTTACGTT TAGCGAAAAT TTGTCTACCC TTGGAAGATA 4920

AAAGTGCCTG ATATTCTTTA GTTTTTAAAC ACTGGTAACG TTCATTCATA TACAGTCTCT 4980

TTTGAGGGGC TGATTCAGGT TCATAATCGC AGTCAACATT GATTTCAAGG CTGTTTGCTT 5040

TCTATCTCCC CGG 5053

(2) INFORMATION FOR SEQ ID NO: 188:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6492 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

AATTCTCTTT TTTCCAACAA AATGTATGAC CTGCACTTGA ATACTTCTCA TTGTTTGTAC 60

ATTCATCTAC TTTCATATAA TCTTTTACAA AATCATAATA TGACATAACA CACTATCCCT 120

TTTAGACAAT ATTCCAATTA GCCTTATTAA TTCAAAACTA TTGTATTAGT AATTATAACA 180

GATGTATAAT AGAAAAGCAA TGATAGATAT TATCAATTAA GCGAATTTAT ATCTAAAAGG 240

GATATTAAAG AAAGGAGATA TGCTTATGAA GATTTACAAA AAACTATTTG CTTATGTCCA 300

AGATAAGAAA TATCTTGGGG TTTTGGCCAT AATTTTTTCT GCTATATCTG CTGCACTTAC 360

AGTATATGGA TATTATTTAA TCTACAAATT TCTAGATAAG TTAATAATTA ATTCAAACTT 420

ATCCGGTGCA GAGAGTATAG CATTAAAATC TGTTATTACA CTAACAAGTG GAGCGATATT 480

TTATTTTGTC TCAGGAATGT TTTCACATAT CTTGGGATTC AGGCTTGAAA CAAATTTAAG 540

AAAAAGGGaA TCGATGGTCT GGAAAAAGCA AGTTTTAGGT TCTTTGACTT AAATCCATCT 600

GGTCAAATAA GAAAGATTAT AGATGACAAT GCTGCACAAA CTCATCAGGT GGTAGCACAC 660

ATGATTCCCG ATAGTTCTCA GGCAATAATC ACACCCGTAC TTGTACTTGC ACTTGGCTTT 720

ATAGTAAGTA TAAGAGTTGG CATAATTTTG CTTGCTCTTA CTATAATTGG TGGCTTAATT 780

TTAGGGGCAA TGATGGGCGA GCAAGAATTT ATGAAGATAT ACCAAGAATC CCTATCTAAA 840

CTAAGTGCTG AAACTGTTGA GTACGTGAGA GGAATGCAAG TTGTAAAAAT ATTTAAAGCA 900

AATGTAGAGT CTTTTAAAAG CTTTTATAAG GCGATAAAAG ATTACTCAAA GTATGCTTAT 960

GATTATTCCC TATCTTGTAA AAGGCCTTAT GTTTTGTATC AATGGTTATT TTTTGGACTG 1020

ATTGCAATTT TAATTATTCC TATAGTTTAT TTTATGACTA GCTTAGCTAG CGCAAAGGTG 1080

ATTTTACTTG AGCTTATCAT GATTTTATTT TTATCAGGAG TTCTCTTTGT TTCATTCATG 1140

AGAATGATGT GtACTCCATG TATATTTCTC AAGGAAATTA TGCAGTAGAT ACTTTAGAGG 1200

CGCTTTACGA AGATATGCAA AAAGACAAAT TAGTGCATGG TAATGTCAAT AATTTTAAAA 1260

ACTATAATAT AGAATTTGAG AATGTTAGCT TTGCTTATAA TGATAAAGCT GTCATTGAAA 1320

ATTTATCCTT TAATTTAGAA GAAGGAAAGT CCTACGCACT TGTCGGTTCA TCTGGATCAG 1380

GCAAATCAAC AGTAGCAAAA CTTATATCAG GTTTTTACAA TGTTAATAAA GGAAGCATAA 1440

AGATAGGCGG GATAGCAATA AGTGAATATT CTGACGAAGC CTTAATTAAA GCCATTTCCT 1500

TTGTTTTTCA AGATTCAAAA TTATTCAAGA AGAGCATTTA TGATAATGTA GCGTTAGCTA 1560

ATAAAGATGC GACGAAAGAT GACGTTATGA GAGCCTTAAA ATTAGCAGGA TGCGATTTAA 1620

TATTAGACAA ATTCCCAGAA AGAGAAAATA CAATCATAGG CTCAAAAGGT GTTTATTTAT 1680

CCGGTGGAGA AAAACAAAGA ATTGCAATTG CTAGAGCAAT TTTAAAGGAT TCCAAAATTA 1740

TTATTATGGA TGAAGCATCA GCATCTATTG ACCCAGATAA CGAGTTTGAA TTGCAAAAAG 1800

CTTTTAAAAA TCTTATGAAG GATAAAACAG TTATCATGAT TGCACACAGG CTATCTACAA 1860

TTAAAGACCT TGATGAAATT ATTGTCATGG ATAGTGGAAA AATTATAGAA AGAGGGTCTG 1920

ACAAAGAATT AATGTCAAAA GATACAAGGT ATAAGAGCCT GCAAGAGATG TTTAACAGTG 1980

CGAATGAATG GAGGGTTTCA AATGAAAGAG TTTTATAAAA AAAGATTTGC TCTTACAGAT 2040

GGAGGAGCAA GAAATTTAAG TAAAGCAACA CTGGCTTCAT TTTTCGTTTA TTGTATAAAC 2100

ATGCTTCCTG CCATATTACT TATGATTTTT GCTCAGGAAG TTTTGGAAAA TATGGGCAAA 2160

AGCAATGGCT TTTATATAGT ATTCTCAGTT TTGATTTTGA TAGCAATGTA TATTTTGCTT 2220

TCTATCGAAT ACGATAAATT ATATAACACA ACCTATCAAG AAAGTGCAGA TTTAAGAATA 2280

AGGACAGCGG AGAATTTATC AAAATTACCT CTATCTTACT TTTCTAAACA TGACATTTCC 2340

GACATTTCAC AAACAATCAT GGCTGATATT GAAGGCATAG AGCATGCAAT GAGCCACTCA 2400

ATACCAAAGG TGGGCGGCAT GGTACTGTTT TTCCCATTAA TATCTGTAAT GATGCTAGCG 2460

GGCAATGTCA AGATGGGTTT AGCTGTAATT ATTCCATCTA TTTTAAGCTT TATATTTATA 2520

CCTTTATCTA AAAAATATCA GGTTAATGGA CAGAATAGAT ATTATGATGT CTTAAGAAAA 2580

AACTCAGAAA GCTTTCAAGA AAATATCGAA ATGCAAATGG AGATTAAAGC ATATAATTTA 2640

TCGAAGGATA TTAAAGATGA CTTATATAAA AAAATGGAAG ATAGTGAGAA AGTACACTTA 2700

AAGGCGGAAG TAACTACAAT TTTAACTTTG TCTATATCTT CAATATTTAG CTTTATATCT 2760

CTTGCTGTTG TGATATTTGT CGGCGTAAAT CTAATTATTA ATAAAGAGAT AAATTCTCTC 2820

TACCTTATAG GATATTTACT AGCTGCTATG AAGATAACAG ACTCTTTAGA TGCATCTAAA 2880

GAGGGCTTGA TGGAAATATT TTATTTATCG CCCAAAATAG AAAGATTAAA AGAAATTCAA 2940

AATCAAGATT TACAAGAAGG CGATGACTAT AGCTTAAAAA AATTTGATAT TGATCTAAAA 3000

GATGTTGAGT TTGCCTACAA TAAAGACGCA AAAGTTTTAA ATGGTGTAAG TTTTAAAGCT 3060

AAGCAGGGAG AGGTCACTGC TTTGGTAGGT GCAAGTGGCT GCGGTAAAAC AACTATCTTG 3120

AAACTTATAT CAAGACTTTA TGATTATGAC AAGGGACAAA TCTTAATCGA TGGCAAAGAT 3180

ATAAAGGAAA TATCAACAGA ATCCCTTTTT GATAAGGTGT CTATTGTTTT CCAAGATGTG 3240

GTTCTCTTTA ATCAAAGCGT TATGGAAAAT ATTAGAATCG GTAAGCAAGA TGCAAGTGAC 3300

GAAGAGGTTA AAAGAGCAGC AAAACTTGCA AATTGCACAG ATTTTATAGA AAAAATGGAT 3360

AAAGGTTTCG ATACAGTTAT TGGTGAAAAC GGAGCTGAGC TATCAGGAGG AGAAAGACAA 3420

AGATTATCAA TAGCCAGAGC CTTCTTAAAA GATGCGCCGA TATTGATCTT AGATGAGATA 3480

ACAGCAAGCC TTGATGTTAA CAACGAGAAA AAGATTCAAG AGTCTTTAAA TAATTTAGTT 3540

AAAGATAAAA CTGTTGTAAT CATTTCACAT AGAATGAAAT CCATAGAAAA TGCAGACAAG 3600

ATAGTAGTTC TTCAAAACGG AAGAGTAGAA AGCGAAGGTA AGCATGAAGA GCTTTTACAA 3660

AAATCAAAAA TTTACAAAAA TTTAATAGAA AAGACAAAAA TGGCAGAAGA ATTTATTTAT 3720

TAGGAGGACT ACAATGGATA ATAAAAAATT AAAAGTAAAA GATTTAGTAA GCATCGGTGT 3780

TTTTGGCGTA ATTTATTTTG CCTTCATGTT TGGAGTTGGT ATGATGGGCT TGATTCCAAT 3840

ATTGTTCTTA ATATACCCGA CAGTATTAGC CATAGTTGCA GGAACTGTTG TTATGTTATT 3900

TATGGCTAAG GTTCAAAAGC CATGGGCACT ATTTATATTT GGTATGATAT CACCACTTGT 3960

GATGTTTGCA GCTGGTCATA CCTACGTAGT TGTGGTTTTA TCACTTATAG TAATGATAAT 4020

AGCAGAATTA ATTAGAAAGA TTGGTAATTA TAATTCATTT AAATACAATA TGCTTTCTTA 4080

TGCAATCTTC AGCACATgGA TATGTAGCTC TTTAATGCAA ATGCTTTTAG CAAAAGAAAA 4140

ATATATGGAG TGGTCTTTGA TGACTATGGG AAAAGATTAT GTTGATGTAT TAGAAAAGTT 4200

AATAACTTAT CCTCACATGG CTTTAGTAGC CTTAGGTGCT TTCTTAGGAG GAATTCTTGG 4260

AGCATATATA GGCAAGGCTC TATTGAAAAA ACACTTTTCA AATGGATTAT ATTGTGTGGG 4320

ATACTTTACT CCTTGCCTAA TTTTATGGTG CTATCTGAAT TAAACCCTAT AGTTAAGATG 4380

TTTTTGAGTA TACCTATTGT TATTAGAATG TTTATTTTAC CATTTATGGC AGCAAGCTTT 4440

ATGATAAAGA CCTCGGATGT AGGCGCAATA ATTTCATCGA TGGATAAGCT TAAGATTTCA 4500

AAGAATGTAT CCATACCTAT TGCGGTTATG TTTAGATTCT TCCCATCTTT TAAGGAGGAG 4560

AAGAAAAACA TCAAAATGGC TATGAGAGTA AGAGGGATAA ATTTTAAAAA CCCAGTCAAA 4620

TATCTTGAAT ATGTTTCTGT GCCACTACTC ATTATATCAT CTAATATATC AGATGACATT 4680

GCAAAAGCGG CAGAAACAAA GGCAATAGAA AATCCAATTG CCAAGACCAG ATACATTCGC 4740

GTAAAGATAC AGCTAATTGA TTTTGTTTAT GTTTTAGCGG TTGCTGGACT TATTGTGGGA 4800

GGCTTAATAT GGTTGAAATA AAAAATTTAA GTCTTGATTA TGGTGAAGAG CATATATTAG 4860

ATGATATATC ACTATCCATA GCCGAGGGAG AGTGCGTGCT ATTTACAGGA AAAAGTGGAA 4920

ATGGTAAGTC ATCTTTAATA AATTCAATCA ATGGACTAGC TGTAAGGTAT GATAACGCAA 4980

AGACAAAGGG CGAAATAATT ATTGATGGTA AGAATATAAA AAATTTGGAA CTTTATCAAA 5040

TCTCAATGCT TGTTTCAACT GTTTTTCAAA ATCCTAAGAC ATATTTTTTT AATGTCAATA 5100

CGACATTAGA ATTATTATTT TATTTGGAAA ATATCGGTCT TGCAAGAGAA GAGATGGACA 5160

GGCGTTTGAA GGATATACTT GAGATATTCC CGATAAAAAA TCTTTTGAAC AGAAATATAT 5220

TTAATCTATC CGGCGGTGAA AAACAAATTC TTTGCATTGC AGCTTCTTAT ATAGCAGGTA 5280

CAAAGATTAT AGTTATGGAT GAGCCTTCAT CGAATTTAGA TATTAAAAGC ATAAGTGTTT 5340

TGGCAAAGAT GCTAAAGATA TTAAAAGAGA AAGGCATAAG CATAATTGTT GCAGAGCATA 5400

GAATTTATTA TTTGATGGAC ATAGTTGACC GTGTATTTTT AATAGATAAA GGAAAGCTTA 5460

AAAAAACTTA TACTAGAAGT GAATTTTTAA AGCTAGATAA AAATGAATTA AATGCTTTAA 5520

GTTTAAGAGA TAAAGAATTA AGTAAATTAA AAGTTCCTTA TTTAAAAGAA GGTGGAGAGT 5580

ATCAGATAAA AAATCTTAGT TACAAATTTA CTGATGATGA GTGTTTAAGC TTAAAAGATA 5640

TTTCGTTCAA GCTTGGGAAA ATTTATGGCA TAATAGGATC CAACGGACGA GGAAAATCAA 5700

CGCTTTTAAG ATGTTTAATA GGTCTTGAGA AAAAATCAAA AGAAGAAATT TATTTTAAGG 5760

GAGAGAAGCT ATCTAAAAAA GAAAGACTCA AAAACTCTTC ACTTGTTATG CAAGATGTAA 5820

ATCATCAATT ATTCACAGAT GAAGTATTCA ACGAGCTTAG ATTAGGAGTA AAGAATTTTG 5880

ATGAAGAAAA GGCGAAAATC ATTTTAAACC CCAATTATTC ACCCCAAATC TAAAAACCAT 5940

CCAGAATCCT TGCCTTAGCT TAGATCCTGG ATGGTTTCTT TTTTCACCCA ATGGGTGTTT 6000

TTTACTAGAC AAAAAAGAGT TTCCCCTTTA TGGTATAAGT GTAGAAAAAA ACACAAAAAG 6060

AAAGGAAACT CACATGAACA GTTTACCAAA TCATCACTTC CAAAACAAGT CTTTTTACCA 6120

ACTATCTTTC GATGGAGGTC ATTTAACCCA GTATGGTGGT CTTATCTTTT TTCAGGAACT 6180

TTTTTCCCAG TTGAAACTAA AAGAGCGGAT TTCTAAGTAT TTAGTAACGA ATGACCAACG 6240

CCGCTACTGT CGTTATTCGG ATTCAGATAT CCTTGTCCAG TTCCTCTTTC AACTGTTAAC 6300

AGGTTATGGA ACGGACTATG CTTGTAAAGA ATTGTCAGCT GATGCCTACT TTCCAAAATT 6360

GTTGGAAGGA GGGCAGCTTG TTCACAGCCA ACCTTATCCC GTTTTCTTTC CAGAACTGAC 6420

GAGGAAACAG TCCATAGTTT GCGATGCCTC AACCTTGAAT TGGTCGAATT CTTTTTACAT 6480

GTTCACCAGC TG 6492

(2) INFORMATION FOR SEQ ID NO: 189:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7174 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

AACTGAAGGT AAAGGCTTCG ACGCAGAACG TGACGCTGCC CAAGCTGCCC TTGATGACCT 60

TAAGAAAGCT CAAGAAGACA ACAACTTGGA CGACATGAAA ACAAAACTTG AAGCATTGAA 120

CGAAAAAGCT CAAGGACTTG CTGTTAAACT CTACGAACAA GCCGCAGCAG CGCAACAAGC 180

TCAAGAAGGA GCAGAAGGCG CACAAGCAAC AGGGAACGCA GGCGATGACG TCGTAGACGG 240

AGAGTTTACG GAAAAGTAAG ATGAGTGTAT TGGATGAAGA GTATCTAAAA AATACACGAA 300

AAGTTTATAA TGATTTTTGT AATCAAGCTG ATAACTATAG AACATCAAAA GATTTTATTG 360

ATAATATTCC AATAGAATAT TTAGCTAGAT ATAGAGAATT ATATTAGCTG AACATGATAG 420

TTGTATCAAA AATGATGAAG CGGTAAGGAA TTTTGTTACC TCAGTATTGT TGTCTGCATT 480

TGTATCGGCG ATGGTACCAG CTATGATATC ATTAGAAATA CAAACATATA AATTTGTAAT 540

ACCGTTCATA ATTGGTATGA TTTGGACAGT AGTTGTATTT CTTATGATCA ATTGGAATTA 600

TATAGGCAAA TACTAAGAAG AGACAAAAAT ATATAAATAT TTCTGTACTT ATAGGATATT 660

TAAAATCAAA ATAAAGTTAA TTTACTTATT TGCAGAGGTT GCAACCCAGC CTCTGTTTTT 720

CGATAAAAAG GGACGGAATC TCATTTGTTT GGGTTTTGTC TCATCAATAG AAAGGAACAA 780

AGAGTGTTCG TAACTGAACA CGGGTTTCAG AATTTCTTAC TAAATATAAA AGAAAGGAAT 840

TGAACCCGAC CTAAATGGTG GTTCGATTCA GAACATCAAT AGAAAGGAAT AAGGGTGTTC 900

GTAACTGAAC ACGGGCTATG GACTGTGCCA AAAAGATAGT TTTTTCTAGG ACGTAAGCGT 960

CCGTCGTCAA AACTCCTAGA TGGCTGTGTC CGTTTGACGC CCTTTGTATC TTGAATTATG 1020

AACAATACTG AATTTTATGA TCGTCTGGGG GTATCCAAAA ACGCTTCGGC AGACGAAATC 1080

AAAAAGGCTT ATCGTAAGCT TTCCAAAAAA TATCACCCAG ATATCAACAA GGAGCCTGGT 1140

GCTGAGGACA AGTACAAGGA AGTTCAAGAA GCCTATGAGA CTTTGAGTGA CGACCAAAAA 1200

CGTGCTGCCT ATGACCAGTA TGGTGCTGCA GGCGCCAATG GTGGTTTTGG TGGAGCTGGT 1260

GGTTTCGGCG GTTTCAATGG GGCAGGTGGC TTCGGTGGTT TTGAGGATAT TTTCTCAAGT 1320

TTCTTCGGCG GAGGCGGTTC TTCGCGCAAT CCAAACGCTC CTCGCCAAGG AGATGATCTC 1380

CAGTATCGTG TCAATTTGAC CTTTGAAGAA GCTATCTTCG GAACTGAGAA GGAAGTTAAG 1440

TATCATCGTG AAGCTGGCTG TCGTACATGT AATGGATCTG GTGCTAAGCC AGGGACAAGT 1500

CCAGTCACTT GTGGACGCTG TCATGGCGCT GGTGTCATTA ACGTCGATAC GCAGACTCCT 1560

CTTGGTATGA TGCGTCGCCA AGTAACCTGT GATGTCTGTC ACGGTCGAGG AAAAGAAATC 1620

AAATATCCAT GTACAACCTG TCATGGAACA GGTCATGAGA AACAAGCTCA TAGCGTACAT 1680

GTGAAAATCC CTGCTGGTGT GGAAACAGGT CAACAAATTC GCCTCGCTGG TCAAGGTGAA 1740

GCAGGCTTTA ACGGTGGACC TTATGGTGAC TTGTATGTAG TAGTTTCTGT GGAAGCTAGC 1800

GACAAGTTTG AACGTGAAGG AACGACTATC TTCTACAATC TCAACCTCAA CTTTGTCCAA 1860

GCGGCTCTTG GTGATACAGT AGATATTCCA ACTGTTCACG GTGATGTTGA ATTGGTTATT 1920

CCAGAGGGAA CTCAGACTGG TAAGAAGTTC CGCCTACGTA GTAAGGGGGC ACCGAGCCTT 1980

CGTGGCGGTG CAGTTGGTGA CCAATACGTT ACTGTTAATG TCGTAACACC GACAGGCTTG 2040

AACGACCGCC AAAAAGTAGC CTTGAAAGAA TTCGCGGCTG CTGGTGACTT GAAAGTAAAT 2100

CCAAAGAAAA AAGGCTTCTT TGACCATATT AAAGATGCCT TTGATGGAGA ATAATACTCT 2160

TCGAAAATCT CTTCAAACCA CGTCAGCGTT GCCTTGCCGT ATATATGTGA CTGACTTCGT 2220

CAGTCGTATC TACAACCTCA AAACAGTGTT TTGAGCAGCC CGTGGCTAGT TTCCTAGTTT 2280

GCTTTTTACT TTATAGATTT TTTAAGACTT TCCTAAGTAA TGACGGACGG TAGTGACCTC 2340

CTTCGAAGTT CCATACCTAA ACTTTGAACC TAAGTTTTAA AGTTTCCGGA CAGCTGAAAC 2400

CAAGCTGTTT CAGGTGTTTT CATTACGGCA GAAAGTCTTC GATTTAGTTG TGAAATGGTG 2460

AATGATACTC TTCAAAAATT TCTTCAAACC ACGTCAGCGT CGGCTTGTCA TGGGTATGGT 2520

TACTGACTTC GTCAGTTCTA TCCACAACCT CAAAACAGTG TTTGAGCTGA CTTCGTCAGT 2580

TCTATCCACA ACCTTAAAAC GGTGTTTTGA GCAGTCTGTG CCTAGCTTTC TAGTTTGCTT 2640

TTTGATTTTT ATTGAGTATG AATTACCTAA ATTATGATGC ATAGTTGATG GGATATATAT 2700

AATAGATTGA AATAGAATAT GAACAAATTG ATAAGAGGAT TTTAAAGTAA TCTCTAACAA 2760

TGCTTTAGAA ACTATGGTGT GCTATTCTAA ATTCAATTCA CTATAACTTG TTTACGTTTT 2820

AAAAAAGAGC CGTCGGGCTC TTTTTACTTA TCTTCAGTTC CCTGCATTTC TTTTATCACA 2880

GCTAGTCTAG TCTGGATATC CTTTTCCAAG ACCTTAAACT TGTAAGTCAA GTCTTCTTGG 2940

TATTCCTTGA TAAGTTCTTT TTGCTGGTTA ATGATTTGCA GGCTGTTTTG GATAATATCC 3000

ACATCGTCCT TGATAGCTTG AACGCGGTCA GTGGTATTCA AGACTTCATC TGTGATGGTT 3060

TGGCGATTTT TTGTAACCAG ATAACTTCCG GCTGCAGCTC CTGCAAATAG CAGTAGGTTG 3120

GATAATTTCA TAGCAACTCC TTAAGCGTTT TTGATGGTTT CAGCGACTTG AGCAAGTTTG 3180

TCAAAGTCTG GTTCGTGGGC GATAAAATCA ATCTTGAGGT CATCGTCAGC ACTGTAGCGA 3240

GGCACAAGGT GAACGTGAGT ATGAAAAACT GTTTGACCAG CGACTTCTTC ACAGTTGGAA 3300

ATGATATTCA TACCAGCAGC CTTAGTGACT TTCATGACTT TTTGAGCTAC TTTTGGTACT 3360

TGGGCAAAGA GTTGGcTGGC GCTCGTAGCA TCCATCTCCA AAAGATTGCG ATAGTGTTCT 3420

TTTGGCACGA CCAAGGTGTG TCCTAGTGTT ACTTGAGAGA TATCAAGAAA GGCAAGGACC 3480

TGCTCATCTT CATATACTTT TGAAGCAGGA ATTTCCCCTG CGATGATTTT ACAAAAAATG 3540

CAATCTGACA TAAAATCTAC CTCTACTGTA CTGAATTTTG ATATAATATA GCTACATTAT 3600

ACCAGATTTG GAGAAAATAT GTTAGAAATT AAAAACCTGA CAGGTGGCTA TGTTCATGTT 3660

CCTGTTTTGA AAGATGTGTC CTTTACTGTT GAAAGTGGGC AGTTGGTCGG TTTGATTGGT 3720

CTCAATGGTG CTGGGAAATC AACGACGATC AATGAGATTA TCGGTCTGTT GGCACCTTAT 3780

AGTGGCTCCA TCAATATCAA TGGCCTGACT CTGCAAGGAG ATGCGACTAG CTACCGCAAG 3840

CAGATTGGCT ACATTCCTGA GACGCCTAGT CTGTATGAGG AATTGACCCT CAGAGAGCAT 3900

ATCGAAACGG TTGCTATGGC TTACGGTATT GAGCAAAAAG TGGCTTTCGA ACGAGTAGAG 3960

CCCTTGTTAA AAATGTTCCG TTTGGAACAG AAATTAGACT GGTTCCCTGT TCATTTTTCA 4020

AAAGGGATGA AGCAGAAGGT CATGATTATC TGTGCTTTTG TGGTGGATCC AAGTCTTTTC 4080

ATCGTGGATG AGCCTTTCCT TGGTCTTGAT CCGCTGGCTA TTTCTGATTT GATTCAGCTT 4140

TTGGAAGTGG AGAAGCAAAA GGGCAAGTCT ATTCTCATGA GTACCCACGT GCTGGATTCG 4200

GCGGAGAAGA TGTGTGATGC CTTTGTCATT CTTCACAAGG GAGAGGTGCG TTCCAAAGGC 4260

AATCTCCTGC AACTACGTGA AGCCTTTGAT ATGCCTGAGG CTAGTTTGAA TGATATTTAC 4320

TTGGCTCTGA CCAAAGAGGA GGATCTATGA AAGACTTGTT TTTAAAGAGA AAGCAGGCCT 4380

TTCGTAAGGA GTGTCTTGGT TATCTGCGCT ATGTGCTCAA TGACCACTTT GTCTTGTTCC 4440

TGCTTGTCCT GTTGGGCTTT CTAGCCTACC AGTACAGTCA ACTCTTACAA CATTTTCCTG 4500

AAAATCATTG GCCTATCCTT TTGTTTGTAG GAATTACGTC TGTTTTACTT TTACTTTGGG 4560

GAGGAACTGC CACCTATATG GAGGCTCCAG ACAAGCTCTT TCTCTTAGTT GGAGAAGAGG 4620

AAATTAAGCT CCATCTCAAG CGTCAAACTG GCATTTCCCT AGTCTTTTGG CTCTTTGTAC 4680

AGACCCTTTT CTTGCTGTTA TTTGCGCCTT TATTTTTAGC AATGGGTTAT GGCTTGCCAG 4740

TTTTTCTGCT CTATGTGCTT TTATTGGGGG TAGGAAAATA TTTCCACTTT TGTCAAAAGG 4800

CCAGCAAATT TTTCACTGAA ACTGGACTGG ACTGGGACTA TGTTATTTCT CAAGAAAGCA 4860

AGCGTAAGCA AGTCTTGCTT CGTTTCTTTG CCCTCTTTAC GCAGGTCAAG GGAATTTCAA 4920

ACAGCGTTAA GCGTCGTGCC TATCTGGACT TTATTTTAAA GGCTGTTCAG AAGGTGCCTG 4980

GGAAGATTTG GCAAAATCTC TATCTGCGTT CTTATCTGCG AAATGGCGAC CTCTTTGCTC 5040

TCAGTCTTCG TCTTCTCTTG CTTTCCTTGC TGGCGCAGGT TTTTATCGAG CAAGCTTGGA 5100

TTGCGACAGC AGTGGTAGTT CTCTTTAACT ACCTCTTGCT CTTCCAGTTG CTGGCCCTCT 5160

ATCATGCCTT TGACTACCAG TATTTGACCC AACTCTTTCC GCTGGACAAG GGGCAAAAGG 5220

AAAAAGGCTT ACAGGAGGTA GTTCGAGGAT TGACCAGTTT TGTTTTACTT GTGGAATTAG 5280

TTGTTGGGTT GATTACCTTC CAAGAAAAAC TAGCCCTTCT AGCCTTACTA GGAGCTGGTT 5340

TGGTTTTACT AGTCTTGTAT TTGCCTTATC AGGTAAAACG TCAGATGCAG GACTAACATT 5400

GCTGATACGA CACTAAAAAA GAAGTTGAGT TCAGTCTGTC TCAACTTCTT TTTTGTTACT 5460

ACAGGATAAT GGTTGGTCCG TAGAGACTTA TACTCTTCGA AAATCTCTTC AAACCACGTC 5520

AGCGTCGTCT TACCGTACTC AAGTACAGCT TGCGGCTAGC TTCCTAGTTT GCTCTTTGAT 5580

TTTCATTGAG TATTAACTTG GTCTTGACTT GGTCAAAGTG GAAGCGGTCA TAGGCCCGCC 5640

AAGCGGCGCG AGTTGGAGCA TCTGGATCAA GAGCGCTGAG TCCCATGAGA AGACTGGAAG 5700

TCTGGTAAAA TTTTTCTAGT TCAATCAAGA ATCGATTATC CACTGTTTCA GCCTTGGCTA 5760

GAAAACCAAG AATAGAGTTT AATTGCTCCT GAAAGCGGAC GTCGTCAGCG CTTGCCTGTT 5820

TGCATGCTTG GTAGGCTTTG TTTAAGTCAG TAATCAAAGT ATGAGCTCTT TTGATGGGGT 5880

CTGTATCTGT CATGGGAATG CCTCCTTTAA TCTGGGTGCC AGTCTTACTT CTGGCAACTG 5940

TGTTTTGATA CTGTTAGTTT ATCACTTTTA ATTCTTTTTT TTTATTCAAA TCTTTAATTG 6000

TCATTGAAAT GTCTTGAATT GCGCTGAGTG AATTTTATGA TAAAATAGTT GTAAGCTCAT 6060

CATGATGTTG TAGAAAATAA TCCTTTTAGG AGTTTTCAAA GACTGTTTAG GATTGGGTGT 6120

GCTTGGGCTA GACCTTTTCT GTTATTCTTT TCTTAGGAGG AGAATCCAAT GAAATATATG 6180

ATTATTCAGA CGCAGAAAAC AGTCTATAAA GTAAACATCG ACGATATCTA CTATATCCAA 6240

ACACATCCAA CTAAAGCCCA TACCGTACAG ATTGTTACAG AAGAAGCTAG TTTTAATATG 6300

CTTCAAAATT TAAGTAATCT TGAGAACCAA TGTGGGGAAA CCTTGATGAG ATGTCATCGA 6360

AATTGTTTGG TTAATCTTGA TAAATTAAAA TCGATTGATT TTCAAGAAAG AATCCTTTTT 6420

CTCGGAGAAG AAGGTCAATA CGCTGTCAAG TATGCCAGAC GTCGCTATAG AGAAATTCGT 6480

CAAAAATGGT TGAAAGAGGG AGAGTAAGAA GATGAGAATA TTTGTTTTAG AGGATGATTT 6540

TTCCCAACAG ACTAGAATTG AAACGACGAT TGAGAAACTT TTGAAAGCAC ATCATATCAT 6600

TCCTAGCTCT TTTGAGGTAT TTGGCAAGCC GGACCAACTG CTGGCTGAAG TGCATGAGAA 6660

GGGGGCCCAT CAGCTATTCT TTTTGGATAT TGAGATTCGA AATGAAGAGA TGAAGGGACT 6720

GGAAGTGGCT AGAAAGATTC GGGATCGGGA TCCTTATGCC CTGATTGTCT TTGTGACGAC 6780

TCACTCGGAG TTTATGCCCC TGTCTTTTCG CTACCAAGTG TCTGCTTTGG ACTACATTGA 6840

TAAGGCCTTG TCAGCAGAGG AGTTTGAATC TCGGATCGAG ACAGCCCTCC TCTATGCCAA 6900

TAGTCAAGAT AGTAAAAGTC TGGCGGAAGA TTGCTTTTAC TTTAAATCAA AATTTGCCCA 6960

ATTTCAGTAT CCTTTTAAAG AGGTTTACTA TCTCGAAACG TCGCCCAGAG CCCATCGTGT 7020

TATTCTCTAT ACCAAGACAG ACAGGCTGGA ATTTACAGCG AGTTTAGAGG AGGTTTTCAA 7080

GCAGGAGCCC CGTCTCTTGC AGTGCCACCG CTCTTTTCTC ATCAATCCTG CAAATGTGGT 7140

GCATTTGGAT AAGAAAGAAA AACTGCTTTT CTTT 7174

(2) INFORMATION FOR SEQ ID NO: 190:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3207 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

CCACCAGGGA AAATCATTGA AGTTGGTAGT CACCAAGAGT TAATGCAGGC GCAAAGTTTC 60

TACCATCATC TATTCAATAA ATAAGGAGAA TGTCATGAAT CCTAATCTTT TTAGAAGCGT 120

CGAGTTTTAT CAGAGACGTT ACCATAACTA TGCGACAGTG TTAATTATAC CTCTTTCATT 180

ACTATTTACT TTCATCTTGA TTTTCTCCCT TGTTGCCACA AAAGAAATTA CTGTTACTTC 240

CCAAGGAGAA ATCGCCCCTA CAgTGTCATT GCCTCCATTC AGTCAACCAG TGATAATCCT 300

ATCCTAGCTA ATCATTTAGT GGCAAATCAA GTAGTTGAAA AAGGGGACTT ACTCATCAAA 360

TACTCTGAAA CAATGGAAGA AAGTCAGAAA ACTGCCTTAG CAACTCAATT ACAAAGACTT 420

GAGAAGCAAA AAGAAGGACT TGGAATTTTG AAACAAAGCT TAGAAAAAGC GACTGATCTT 480

TTTTCTGGCG AGGATGAATT TGGCTACCAT AATACCTTTA TGAATTTTAC TAAACAATCC 540

CATGATATTG AACTGGGTAT CACAAAGACT AACACCGAAG TTTCAAATCA AGCTAATCTT 600

TCCAATAGCA GTTCATCAGC TATTGAACAA GAAATTACAA AAGTTCAACA ACAAATTGGA 660

GAATATCAAG AGTTGAGAGA TGCTATCATA AATAACAGAG CACGCTTACC AACTGGCAAT 720

CCGCACCAGT CAATTTTGAA TCGTTATCTT GTAGCCTCAC AAGGACAAAC ACAAGGAACT 780

GCAGAGGAGC CATTTTTATC TCAAATTAAT CAAAGTATTG CAGGTCTTGA ATCATCTATC 840

GCAAGCCTCA AAATTCAGCA AGCTGGTATC GGAAGTGTAG CAACTTATGA TAACAGTTTA 900

GCAACCAAAA TTGAAGTACT CCGCACTCAG TTTTTACAGA CAGCCTCACA GCAACAACTA 960

ACTGTGGAGA ATCAATTAAC AGAATTAAAA GTACAACTAG ATCAAGCCAC ACAGCGTTTG 1020

GAAAACAATA CCTTAACCTC CCCAAGTAAA GGTATCGTTC ATCTGAACAG CGAATTTGAA 1080

GGTAAAAATA GAATTCCAAC TGGTACAGAA ATTGCTCAAA TATTCCCTGT CATCACAGAT 1140

ACAAGAGAAG TACTAATCAC TTACTACGTA TCTTCTGACT ATCTACCTCT ACTAGATAAA 1200

GGACAAACTG TAAGATTAAA ACTGGAGAAG ATTGGAAATC ACGGCACCAC CATCATCGGC 1260

CAACTTCAGA CAATTGATCA AACTCCTACC AGAACAGAGC AAGGAAATCT CTTTAAATTA 1320

1150

ACCGCTCTTG CAAAACTATC TAACGAGGAT AGTAAACTCA TCCAATATGG CTTACAAGGT 1380

CGCGTCACTA GTGTAACTAC AAAGAAAACA TATTTTGATT ATTTCAAAGA TAAAATTTTA 1440

ACACATTCTG ATTAATTTTC AGATAACACT CTATAACTAT TTATTATCTT ATCAAAAAGG 1500

AGAATCATAA CATGGATAAG AAACAAAACC TAACTTCATT TCAAGAACTA ACAACTACCG 1560

AACTCAATCA AATTACAGGT GGAGGATTGT GGGAAGATTT ATTATATAAC ATTAATAGAT 1620

ATGCTCATTA CATCACATAA GAACTTCATC ATCCAATACA ACTATAAAAA AATAAGACCG 1680

AGAAACAAGT ACTCTCGGTC TTATTTTTCA TCATTCTGTA TGTATCACAG TAAGTACCTG 1740

ACGAAAGACT TGATTTTGAC AGGTGGTATT TAGACTGGTA TTAGGATGGC TTTCCACAAT 1800

CTTCATGACG GTATAGAGAC CAACTCCTCT CTCCTCCCCT TTAGAACTGG CTCCAAAGGA 1860

GAAGATTTCA GAAATATCGA TGCCCTCTTC TTTGATGGAG TTTTCGATGA TAAAGGTCTC 1920

CTGTGCTCCA TTTTTTAAAA AGGCGATTGA AACATGAGGT TGACTAGCTT CCACACTGGC 1980

TTCAATAGCA TTGTCACAAA GGATAGACAC AATGGTTAGA AAATCAAGTA GACTCATCCC 2040

CTCGACCTGA ATCTCCTCAG GAACTTCGAC ATTAAAGACA ATGTTCTTAT CTCTGGCTTT 2100

TAAAAATTTC CCTGCTAGAA GACTTTTGAG GGCTTTATCA CGAATATTTA CCAATCTGCC 2160

CAGGTCATAT TTATTGTTCT GCAATTTCTG ACTGGAATCC TTTAAGACGG AGCCATAGAC 2220

CTCTTTTATC TGCTCCATAT CCTCCTCTTC AATGCCCAGA CGTAAGCTAG TCAAGAGGTT 2280

GGTATAATCA TGACGAAAGC TCCGTACTTC CTTGTAAAGC TCCTCTATAT GCCGACTATA 2340

GCGTTCCATA TCTCTATAGC GCAGGGCCTG CTCTTGTTCC AATCTCTCAT AGAGTTTTTC 2400

CTTCAAATAG GTATCCAATT TCTTGATAAC CCCCATAAAA AAGAGTAGGT AAAAGACTAG 2460

GATGAGATGG CGAACAGTCT TTGATTGAAT ACTTTGTTCA TATTCAAAAA AAGACAGACT 2520

TTCCATGACT AGATAGTAGC CACCCATTAT CCAGTTAATC TGAGTCAGGG ACTTTTGAAA 2580

GGCTTTATCG AGAATCTCCT TTCTCAAGCT AGTAAAATCG TAGTCCAACC ATTTCAAAAA 2640

AGCTAGAGAA ATGAAGAAAT TGAAAATTAT TATACATAAC CCAGTAAATG AGTAGCCATC 2700

ATATACTTGC CCTTGTCCCA AAAATGGAAG CACAAAATAG GAGACTCCTC TATAAAAGAG 2760

ATTCACCAAT ATCATTGGAA AGAGACCATA AAAGAAAAGG AGTTTTTTAG GAAGCCCTCT 2820

CAATAATAAG AAAGATAAGC CTATGCCGTA CAAGGGTTCC ATAAAATAAG ATAGGTAAAC 2880

ATTTCCTACT ATATAGCTAA TCATCACAAA AACAAAGGCC AACAGTATCT TCAAAAGAAA 2940

GGCCTTAAAA ATCCTCTCGA AAGTAAGATC AATTCCATCC ACCTTAAAGA AGATGACAAT 3000

TTCTAGTCCA TTAGTAACAA GTGTATACAA CAATATCCAA GCAATGTTCA TAAATTCTCC 3060

TAGCTCAGTG TAATTTATTG ATGGCCTCAG ACACTTCCCT GACCTTATAA CGGGCGATTA 3120

GACAACTTCC ACCATTGGGA GAGAAGAGCA GTTTTTCTTT CTTATCCAAA TGCACCACAT 3180

TTGCAGGATT GATGAGAAAA GAGCGGT 3207

(2) INFORMATION FOR SEQ ID NO: 191:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10357 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

CTGAATCAAG TGTACTGCAC CAGTTCGTGC ATCAGGCATA ACAACATCTA CAGATATAAT 60

ATTGTTTTCT GAGTCCGCCT CATAAGTTAA AATCATAAAT TTTTCGATAT TCGAATTTTT 120

AGTAGCTTGT TCAATTTCTT GAATCATTTC ATCAGAAACT AACTCCATCT GAATTGGAAA 180

GGAATGACTA TTTTCATCAT TTTTGTAGGA AGAATGTTGA TTAAGATAAA GTGTATTCAT 240

CTGAGCATAT TCAAATAAGT AGCCACTCTT ATTTTTTTGT ACCAAAGGAA ATTGGTTTGT 300

AAGTCGCTTC TTACCCTTTA TAATTAACAA TACTTTCCCA TATTTTTCTG TATTTGTTTC 360

AAATTCTAAA TATCCCCAAG TCTGTCCTGC TAATTGTAAT TTATACTCAA ACAAATCTGC 420

TGATGCAAAT GCAGTATCAA TATGATTAGG TCGCGTCCAT GCATAACCAT TCGACACTAT 480

CATTGTCTCT CTTTTTTCTA GACGTTCATC TACATAATCT TTTTGCCCTT TCATCAAAGT 540

ATCTACAATT TTTTGTGCCT CAAGCGAATC AAAGAGATCC TGATTCAACA TAATTCTTCC 600

TCCTCCAAAT ACTTTTTAAT GAATTATACC ATTTTCTTAA AGAAATTACT ACAATAATTA 660

TCTTTTTCTT AAAGTTCTGT GTCAGAGTAA TTTAGAAAAT TATATCTTCT ATAGTAAAAT 720

CAATTAAAAA CTGAACAAAT TTATTGGGAA ATTCAAATCG CTTTCTGAAA ATATTTTAGG 780

AACCGTAGTG TAATATTCCA GATTCAATTC ACTATAAAAC TGACCTTTCT CCTGCAAAAG 840

AAAAAGGAAA GACTTCCTTT CGTGCCTTTC CTCTTACTTG CTACTTGTTT GATTATTTTT 900

GGTAAGCTAC TGCTTGTCTG ATAAAATCCT GAATCGGCTC TCCTTGGTGG AGAGCTTTTA 960

CTATTTTCGA ACCGACGATA ACACCATCTG ACACCGCATT GAAGCGTTCC AGATCGGCTT 1020

GACTAGATAC ACCAAAACCT GTCAAGACTG GGATGTCGGC CACTTGATGA AGTTGCGCCA 1080

AGTGCTTGTC CAAATCTGCA CGGTAATTGC CTGATTTCCC TGTCACTCCA TTGATGGCAA 1140

CGGCATAGAT GAATCCCTCC GCCCCTTCAA TCAACTCTTT CTGGCGCTCA ATTCCTGTGG 1200

TCAAGCTTAC TAAAGGAATC AAGGCGATAT CTGTATTTGC CAAAAATGGT TCTACAAAGT 1260

TGGCATGTTC ATGAGGCAGG TCTGGGATAA TCAAGCCCTT CACAGCTGTA TCAGCCAGAT 1320

CTTTGACAAA GTTCTCCACA CCGTACTGAA AGAGGGGGTT GAAGTAGGTC ATGATGACCA 1380

GTGGAATCTC TGTTTCAATG GTTTTCAAGG TTTCAACTAA AGCCTGGGTA GAGGTCCCGT 1440

GGGCTAAACT GCGCAAGCCA GCTTCTTCGA TAACAGGTCC ATCTGCAACA GGGTCTGAAA 1500

AGGGAATACC CACTTCAATT GCAGAGACAC CCAAATCTTC TAAAAAGTGA ATTGTTTCAG 1560

CAAGACCGTC CAAACCTTTC TCGTGGTCAC CAGCCATGAT ATAGGGAACA AAAATTCCTT 1620

TTCCAGCTGC TTTAATAGCA TTTAATTTTT CTGTTAGTGT CTTAGGCATG AGCTTCTCCC 1680

TTCTTTGCTG CATCTGCTTC CAAGCGGTCC TTGACTTGAA CCACATCCTT GTCCCCACGA 1740

CCTGATAGGC AGACAATCAT AGACTTTTCT GGTCCAAGTT CTTTGGCCAA TTTCACCGCA 1800

AAGGCGATAG CATGGCTAGA TTCCAAGGCT GGGATAATCC CTTCCACACG AGACAAGAGT 1860

TGGAATCCTT CCAAGGCTTC TTCGTCTGTC ACAGGGACAT AGCTGGCACG TTTAATATCG 1920

TGGTAGTGAG AATGCTCTGG ACCGATACCA GGATAGTCCA AACCTGCTGA GATAGAGAAG 1980

GCTTCAAGAA TTTGACCATG GGCATCTTGG AGCACATCCA TGAGGGAACC GTGAAGGACA 2040

CCTGGACGAC CCTTGGTCAA GGTAGCTGCG TGGTGCTCTG TATCCACACC AAGCCCTGCT 2100

GCTTCAGTTC CATACATAGC TACTGACTCA TCTTCTACAA AGGGATGGAA GAGCCCGATA 2160

GCATTCGACC CACCACCAAC ACAGGCTACT AGGGCATCTG GCAGATCTCG ACCTGTCAAG 2220

TCACGGTACT GTTGTTTAGC CTCTCGACCG ATGACACTTT GGAAGTCACG AACGATTTCT 2280

GGAAATGGAT GAGGCCCCAA GGCAGAACCA AGGATATAGT GGGTATCGTC GATATTAGCC 2340

ACCCATGAAC GAAGGGCTGC ATTGACCGCA TCCTTGAGCA CGCGCGAACC ATCTGTTACA 2400

GCCTCGACCT TGGCTCCCAA AAGCTCCATG CGGAAGACAT TGAGGGCTTG GCGTTTGACA 2460

TCTTCCTCAC CCATGTAGAT GGTACATTCC ATGTTAAAGA GGGCTGCAGC AGTTGCAGTT 2520

GCCACACCGT GCTGACCAGC ACCCGTTTCT GCGATAATTT TCTTTTTACC CATGCGTTTG 2580

GCAAGCCAAA CTTGTCCTAA GGCATTGTTA ATCTTGTGGG CTCCTGTATG GTTAAGGTCT 2640

TCCCGTTTGA GATAAATCTT GGCTCCGCCA ATATGCTGGG TCAAGTTTTT TGCGTAATAA 2700

AGAGGAGTTT CACGTCCTAC GTACTGGCGC AAAAGCTGGT TTAATTCCTC TTGGAAACTT 2760

GGGTCTGCCT GACTTTCACG GTAGGCCTTC TCCAACTCCA AAACTGCTGT CATCAATGTT 2820

TCTGGGACAA AACGTCCGCC GAATTTTCCG TAAAATCCAT CTTTATTTGG TTCCTGATAT 2880

GCCATGCTTT ACCCTCTCTA TAAATCTTCT AATCTTTTCA TGATCTTTTT GTCCATCTGT 2940

CTCCACTCCG CTCGATACAT CTACTGCATA GGGAGTAAAG TGTTGAATTG CTTTTACTAC 3000

ATTATCTTCA TTAAGGCCAC CTGCGATAAA GAAGGGCTGT GCTAGTCCAG TCGTATCCAG 3060

TTGACCCCAA TCAAAGGGCT GGCCACTTCC TGCCACAGGG GCATCAAAGA GTAGATAATC 3120

TGCCTGAGAA TTGGGGACAT GCCCATTTCC ATCTACCTGC ACAGCCTGAA TACTGGCACA 3180

AGGCAAATTC TCAAATAAAT CATCTGCCAC CTGACCGTGA ACTTGAACCA AGTCCAAGCC 3240

AACTTTGTCA ATCGCTTCCA GCAGTTCTAC CCGACTTGGT GAAACAAATA CTCCAACCTT 3300

TTTCACATCT GCAGGAATAA GCTTTGCCAA CTCAGCTGCC TCTTCTAAAG TCACCTGTCT 3360

TTTACTAGGT GCAAAGACAA AACCGATATA GTCGGCTCCT GCTGAAACGG CTGTTTCCAC 3420

CGCTTCTTTG GTCGATAGTC CACAAATTTT AACCTTTGTC AATCTGCAAC TCCTTGATTC 3480

TCTGGGCCAC ATTTTCTGCC TGCATAAGAG CTGTCCCTAC CAAAATTCCG TTAAAGTATG 3540

GGGCTAGTCG TTCCGCATCC TGCCCTGTGA AAATGGCAGA TTCAGAAATG TAATAGCGAC 3600

CTTCCTCAAA GTAAGGGGCT AAATCTACAC TGGTCTGCAA GTCGACCTCA AAGGTAGTCA 3660

AGTTGCGGTT GTTGACCCCG ATAATCTCAG CACCAAGTCT GTGGGCTACC TCTAGTTCAG 3720

CTAGATTGTG AGTCTCCACT AAGACTTCCA GACCAAGCTC TGTCGCGTAG TCATACAGTT 3780

CCTTGAGGCG TTCTTCGGAC AAGGCTGCCA CAATGAGCAA GATAACTGTC GCACCTGCAT 3840

TGCGAGCGCG GATGATTTGC TTTTCATCGA TGATAAAGTC TTTGTTGAGC GTCGGAATCT 3900

CTACCTGACT GGAAATTTCC CGTAGATAAT CCAAATGCCC TTTAAAGAAA ACCTCATCTG 3960

TCAACACCGA AATCATCACT GCTCCGTTTT CTTCATAAGT CTGGGCCTGT TGCACAATAT 4020

CCACATCGAG ATTGATATCT CCCAAACTAG GGCTAGCTTT CTTGACCTCA GCGATTACCT 4080

GCAAGCGGTC CTGATGATTC TTCAAAAATT CTGCCAAGCG ATAGGTCTGG CGCAGAGGCT 4140

GGATTTGCTC CAGCTTCATC TGCTCCACCT CACGCGCCTT CTGCTCTAAG ATTCGTGCTA 4200

AAAATTCCTG ACTCATTTTT GGTACTCCTG TAACAGTCTG AGTTTTTCAA GGGCCTTGCC 4260

TCTAGCAATC ACTTGACGGG CCAAGGCAAC CCCTTCCTTG ATGCTATCAA TCTTACCATT 4320

AGCATAGAAA CCAAGACCAG CATTCAAGAC TGTCGTTTCC AAGAATGGAC TTGCTTCGTT 4380

TTTCAGAACG CTAAGCAAAA TTTCTGCATT TTCCTGAGCA TTCCCACCAC GAATATCTTC 4440

CATAGCATAG CCTTCCATTC CCAAATCCTC TGGAGTAAAG CTTGACAAGC TGATTTCGCC 4500

ATTTTCAAGA AGTGCAATCT TGGTTGTTCC GTTCAAGCCA GCTTCATCCA ACCCTTCTGG 4560

TCCAGCAACC ACGATGGCAC GTTTGCGACC CATATTTTTC AAAACCTGAG CTGTACTTTC 4620

TAGGAGTTCT GGACGACTAA TTCCAAGAAG CTGTGTTTCT AAAGCCATTG GATGAATCAG 4680

TGGACCAGTC AAGTTCATAA TCGTTGGAAT TCCCAATTCC AAACGAGCTG GCATGATGTA 4740

TTTCATAGCT GGGTGCATAT TTTTAGCGAA GAGAAAGACG ATTCCAGTTT TATCAAAGAC 4800

CTTACCTAGT TCAGCTGGTT TGAGGTCAAG ATTGATTCCC AAGGCTTCGA GGACATCTGC 4860

GGAACCAGAT TTAGAAGATA TCGAGCGGTT ACCGTGTTTG GCCATGTGAA TACCGCCACC 4920

AGCCAAGACA AAGGCTGCAG TTGTGGAAAT ATTAAAACTG AAAGACTTGT CCCCACCTGT 4980

ACCACAGTTG TCCATGGCAT CATGAATCTC AGTTGGAATA TGCTGGGCAT GTCCTCTCAT 5040

GACTTGGGCA ATGGCTGTGC GTTCTTCAGG TGTTTCCCCC TTCATCTTAA GAGCTAAGAG 5100

GAGAGAAGCA ATCTGCGCTT CAGTTACACG CCCAGTTACG ATACGCTCAA TGACATCCGT 5160

CATTTCCACA CCTGATAAAT TTTCAAATTT TGCTAGTTTT TCAATAATCT CTTTCATCCT 5220

AGTTTCCTCA CTTTACAACC TCCTCGATAA AATTCCGAAT AGAAGACAAG CCGTCTGGCG 5280

TTCCAATGCT CTCTGGATGG TACTGGAAGC CATAAATCGG TAGGTTTTTA TGTTGAATCC 5340

CCATGATGGC TTGGTCATCA GTCGAACGAG CTGTCACTTC AAAGTCTTCT GGCATTTCCT 5400

CAATCAAAAT ACTGTGATAA CGCATGACCG CACGGCCATC CTCAATACCT TGATACAAAA 5460

CAGATGGCGC TTCAAAGTTG ATATTGCTCT GTTTCCCATG CATGACTTTT GGAGCCAAAC 5520

CTAGCTTACC ACCAAAGACT TCTGCAATGG CTTGGTGGCC CAAACAAATC CCAAGAATCG 5580

GCTTCTTGCC TGCAAAATCA CGAATCATGT CTTCCATCTT TCCAGCATCA ACTGGCCAAC 5640

CAGGACCAGG AGAAAAGACC AGACCATCTG CTTTTTCAGC TTCTTCATAC AGCTTGGAAT 5700

CATCATTTCT CAGAACCTGA ACTTCTGCAA AATTCCCAAT GTATTGGGCC AAGTTATAGG 5760

TAAAAGAATC ATAGTTGTCA ATCAATAAAA TCATGGTCTT AGTTCTCCAA TTCTAGTCAT 5820

AGATTTTGCT TTGTTAATGG TTTCTTGGTA TTCGTTTTGG GCGATAGAGT CGTAGACAAT 5880

CCCTGCCCCA GCCTGCACAT AGGCTCTTTG ATTTTTGAGA ATCATGGTTC GGATGGCGAT 5940

GGCCAAATCC ATATCACCCG TCGCAGACAA GTAGCCGATT GCCCCAGCGT ATACTCCCCG 6000

TTTTTCCGTT TCCAGTTCAT AGATACGTCT CATCGCTCGA ATCTTTGGTG CTCCAGAAAC 6060

GGTTCCAGCA GGAAGCGTTG CTTTCAAGGC ATCCATGGCA GTGAGTTCTG GAAGCAAACG 6120

CCCCTTGACT ACGCTGGTCA AATGCATGAC GTAGCGGAAG AGCTCCACTT CCATATACTT 6180

AGTGACTTGG ACACTGGTCG TTTCAGAGAT GCGGCCAATA TCGTTACGCC CCAAGTCTAC 6240

CAACATTCGA TGTTCTGCTG TTTCCTTCTC ATCAGAGAGG AGGTCAGTCG CCAAGGCCTT 6300

GTCTTCTTCA TCCGTAGCCC CTCTTGGTCG CGTCCCTGCA ATCGGATTGG TTGTCACGAT 6360

GCCATTTTTG ACAGAAACCA AACTTTCTGG ACTAGCTCCG ATGATTTGAT AATCCCCAAA 6420

ATCATAGAAA TAAAGGTAAT TAGAAGGATT AGTCACGCGG AGATTTCTGT AGAAGTCAAA 6480

TGGATTTCCA GTAACTTCTG CTGAAAAACG CTGGCTGAGT ACACATTGGA ACATATCTCC 6540

GTTACGAATC AAGTCACGAG CTGTTTCTAC CATTCCCTCA AACTTATGTG GAGCGATATG 6600

CGGTTTGAAG TCTAACGGAG ATAGATCCAA ATCTTCAAAT TCATTTGGAG CAGGAATGCG 6660

TAATTCCTCA AGCACTTGGT TCAAGGATTT TTCCAAGGCC TCTTGACTGC GCTCACTATA 6720

AAGTGCATCC TCTATGACAT GTATCTTCTC CTTCTTGTGG TCAAAGACCA TATAGCTCTC 6780

ATAGACAAAG AAATGCATGT CTGGCGTCCC AATTGTATCC TCAGGGATTT GACCAATTTC 6840

TTCATAAAGC GAAATCATAT CGTAACCCAC AAAACCAATG GCTCCACCAC CAAAAGGTAG 6900

CTCTGAGTGG TGCTGACTCT TATGAATCAC TTCATAAAGG AAATCCAAGG GATCCCGATC 6960

AATCACTTGA CCATTTTGAT AGAGAACCCC ATTTTCAAAC TTAATCTCAA AAACTGGATT 7020

ATAGGCTAGG ATAGAAAAAC GAGCTGTTTC CTTGTCTCTC GGAATACTCT CTAAAATAAC 7080

CTTATGTTGC CCCTTTAAGC GCATATAAGC CAAGATTGGT GATAAGACAT CTCCATGAAT 7140

GATTCGTTCC ATTGTAATTT CCCTTTCAGT TCTACTTCTA GTCCGTGGTG ACTGTATGAA 7200

AAATCCCCAC GCAAAATAAC TTGCGTGAGG ACGAAATTCG CGGTGCCACC TCAATTATAG 7260

GATTTCTCCT ATCTCTCATT CCTGTCTCAG ATATCTCCTG TAACAGGCTG TGCGATAAAG 7320

GGCACTCCCT TGAGAATGAT GTTTTCTTCT CTCGTTTCAG ATGAACCCAA CTTTACAGCT 7380

TTCTCTGCTT GTTTTCAGCA ACCACAAGCT CTCTGTGAGA GAAAGAACTG TAATTTTTCC 7440

ATCTATTATT TTTTAGCTTC TAGTAGTCTG CAATCGCAGC TAGGTCCTTG CCTCCACGAC 7500

CAGAGACATT GATGAAGAGA TGTTCATCTC GGTACACCTT TATACTCTTC GAAAATCTCT 7560

TCAAACCGCG TCAACGTCGC CTTGCCGTAG GTATGGTTAC TGACTTCGTC AGTTCTATCT 7620

GCAACCTCAA AACAGTGTTT TGAGCTGACT TCGTCAGTTC TATCCACAAC CTCAAAACAG 7680

TGTTTTGAGC TGACTTCGTC AGTTCTATCC ACAACCTCAA AACAGTGTTT TGAGCTGACT 7740

TCGTCAGTTC TATCCACAAC CTCAAAACAG TGTTTTGAGC AGCCTGCGGC TAGTTTCCTA 7800

GTTTGCTCTT TGATTTTCAT TGAGTATTAC TAGCTTTTTT CGTATTAGTC CAGCCTTTTT 7860

GTTTGCTTTT AGTAGTAGGC ATGGAGCTGT AGATAGAACT CAAGTTCATC AAAGCGACTT 7920

AAGGCCCTAA TAAAAGATAA ACCAAACGAC GGATAGAAAA AAGCCCACAC ACAGAATATA 7980

CTTCCGTGTG AGGGCGTTGG TAACGCGGTG CCACCTCAAT TATAAAGGGA CTATCCCTTT 8040

ACATCTCTGC CTTGTTTAAC AACAAGCTGC ACTGTAAGGT GTGCGCACCG AATTTTCATT 8100

GTTTCAAATT CATTTTCAAA ATCAGCCCAC TTTCACTACT TCCAACCACC TATTCACAAT 8160

CACCACAGGC TCCCTGAAGA TCAAAAATAG TTACTTTTCT GATTTGTTGA ACTTATTTTA 8220

ATACTTTGTT TTTTCTTTGT CAAGACTTTT TTACGATTTT TTTGAAAATA TCATTCGAAT 8280

ATGACCATGT CTTCCTTAGA TCGAACATGA ACATGTCCCA CTTCTTAGAA ATTGGATCCA 8340

ACTCAATAGA AACTGAATGG AGGCTAAACA GAACTTATTT TAGAACACTC CATCTTTTCC 8400

ACTAGGATTT TCAAGAATTA AACAATACTA GAAACTCTGT CTCCTAACAA ATTTAGGAGA 8460

AACTTCAACA GATGTGACAC TTTCCCCTTT AATAATTGCT AAAACACCTT CTATCATTTC 8520

TTTAGCCAAT TTAACATAAT TGGGAGCAAT TGTAGACAAA GCTGGAGTAT AATACTGAGA 8580

AATAGGAATA TTATCAAATC CAATGATAGA AATATCATCT GGAATAAGAA TTCCTTTCTC 8640

ATAGCACGCA CGAATCAAGC CCTGAACCTT TTCATCTCCT GAAACAAAAA TAATGTCCGG 8700

ATAATTTTGG GTAGTCAAGT GCTGCATTGC ATAAGAATAA ACTGAATCAA TTGTAGATAA 8760

GCCATAAATG ACTTTTAAAT CCATAAAGTA ATTTTTATCA TTCAGAAAAG AACGCACACC 8820

TCTTTCACGA TCCTTATTAA CATGGGATTC TCCTCCCATA AGCAACCACA TATTTTTAAA 8880

TTTTTCTTCA GTTACAGCTT TCATCATATC ATAAGTAGCT TGAAAATTAT TATTAGATAC 8940

ATAGACTACT CCAGACGTTT GAGATTCACC GAAAACAAGA AAAGGCATAT GGTTCTTCTT 9000

TAAATACTGA ATTCTGATAT CATCTACACT TTCATAAAAA ACAATAACAC CATCTACTAG 9060

GCTACCTGTG CTTGATATAA TTGAATTACT AATTGTATCC TCCTCTCCAA AGTACTCAAC 9120

TATAGCATTA ACACCAAATT CTTTACACGT CCGTAACACT TTATCTAACA GCGTATGAAA 9180

CCAAATTAAA GGAAAAGAGT CGATTTTTTT TACAGAAATC AATATATTTA TAGCTTCTTT 9240

TTTAGTTAAA TTTTTTGCAT ACGCATTTGG AATATACGAC AATTCCTCTA TAACTTTTTG 9300

AATCGCTTGA TAAGTTTCTT CTTTAACATT TACTCCACCA TTAATAACTC GTGAAACTGT 9360

TTTTGGAGAA AAACCTGATA AACGTGCAAT ATCATAAATA GTTACCTTTT TCCCATTTAT 9420

ATTTTTCATT TCAGTCCTCC ATTACGAACA TTCTAATATT ACTATACAAT ATTTAATTTT 9480

TTTTAACAAG AGAATTTAGT AAATTATTTA AGATCCACAA ATTCACAAAA TTAATTTTAC 9540

AAATATTCTT CCCCTTCAAA AAAGTTTAAA TTGCATTTCA CACCTTTATT TTTAAGAATG 9600

TTTCCAACTT CACGACAAAT AAATTCATAT GAGAAAAAAC TGCCATAAAA TTGTAGATTA 9660

ACTTTTTCAG TAAAATGTGT AGGATTTATA AAAACATATA ATAGCCTGTC AATGTAACAT 9720

TTTAACATAG AGTTAATTTT TTCTTTAAAG ATAACATTTG TTATCAACTC ATCAGGAGGT 9780

AAATGAAAGG CAAACACCAT TTCACAAATA TCATAAAAAG AAATAAATTT GTATACTTGT 9840

ATCAAACAAT TATTATCAAA ATATTCTATT TTACCTAAAT CAAAATTGAT TTTATAATCT 9900

TTCATAAAAA CCTCTGAGCA AAAATCTACT CAAAAATTAG ATGATTAAAA CATCTAAAAA 9960

GCAAAAGGAC AAAAACATCT GTCCCTTTGT TTACTAAATT TCAGCTAATT TCTTCGACAT 10020

AAATAACACC TACAATATTA GCAATTTCTT CCATCAGTCG AAGATGTTCA AATCTACCTG 10080

ATAATTCCAG AGTAATAAAT GACGCTATTT TTTTGTCCGG AACATCAAAG TATTCAATTC 10140

TGTCAGAATT AACATCTCCA AACGCTGTTC TTGAATCGGT CATTCTGATA CCATTTTCTG 10200

CACAATAAAC CAATACACGA TTATAGGCTT CTGTAGATTT AACCACTATA TACAATTCAA 10260

TCATTTTAGA ACGATTTTGC AGATATTTTT TTAGTGGTTG GAACATGGAT ATCACACCCC 10320

AAACAGAAAT GGCTACTAAA AGAGCTCCCT CATAAGG 10357

(2) INFORMATION FOR SEQ ID NO: 192:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6867 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

CGGGACATTC TCAATCTTCT GTCTTTTGTT TTTCTCTTCT TTCTATGATA CAATGGAAAA 60

AATAAATTCA AAAGGAGTTT TTTTATGACT TATCCAAATC TCTTGGACCG CTTCTTAACC 120

TATGTTAAGG TCAACACGCG CTCTGATGAA CACTCTACTA CTACTCCAAG TACACAGAGT 180

CAGGTTGACT TCGCAACAAA TGTCCTAATT CCTGAAATGA AACGTGTTGG ACTGCAAAAT 240

GTTTACTATC TACCGAATGG TTTTGCTATT GGAACCTTGC CAGCCAACGA TCCGTCTTTA 300

ACACGTAAGA TTGGTTTTAT ATCGCACATG GATACTGCTG ATTTTAATGC TGAAGGAGTC 360

AATCCACAGG TAATTGAAAA CTACGATGGT GGTGTGATTG AACTAGGGAA TTCTGGTTTC 420

AAACTCGATC CAGCTGACTT CAAGAGTCTT GAAAAATATC CAGGACAAAC GCTCATCACA 480

ACAGATGGAA CAACCTTGCT AGGTGCTGAT GACAAGTCAG GAATTGCTGA AATTATGACA 540

GCCATTGAAT ATCTAACTGC TCATCCTGAA ATTAAGCACT GTGAGATTCG TGTTGGTTTT 600

GGTCCAGATG AAGAAATCGG TGTTGGTGCC AATAAATTTG ATGCAGAAGA TTTTGATGTG 660

GATTTTGCCT ACACTGTTGA TGGTGGTCCA CTAGGTGAAC TTCAGTACGA GACTTTCTCA 720

GCCGCTGGTG CTGAATTGCA TTTCCAAGGT CGTAATGTCC ACCCTGGTAC TGCCAAAGGG 780

CAGATGGTCA ATGCCCTTCA GCTAGCAATT GATTTTCATA ATCAACTTCC AGAAAATGAC 840

CGACCTGAGT TAACTGAAGG TTACCAAGGT TTTTACCATC TAATGGATGT GACAGGTAGT 900

GTTGAGGAGG CGCGTGCAAG CTACATCATT CGTGATTTTG AAAAAGATGC CTTTGAAGCG 960

CGTAAAGCAT CCATGCAATC TATCGCTGAT AAGATGAATG AAGAACTTGG GAGCGACCGT 1020

GTCACTCTCA ACTTGACAGA CCAGTACTAC AATATGAAAG AAGTCATTGA AAAAGATATG 1080

ACTCCAATTA CCATTGCTAA AGCCGTTATG GAAGATCTAG GTATCACGCC TATTATCGAA 1140

CCAATCCGGG GTGGAACAGA CGGCTCTAAG ATTTCCTTTA TGGGAATCCC AACTCCGAAT 1200

ATCTTTGCAG GTGGCGAAAA TATGCACGGA CGTTTTGAAT ACGTTAGCCT TCAGACTATG 1260

GAACGTGCAG TTGATACCAT CATTGGCATC GTAGCTTATA AAGGCTAAAA AGACGAGGTA 1320

GCTCAGCTAC TTCGCCTTTC TTTTTATTCT ACTGGTTTTT CTTGATTTCC AGTAGTTGTA 1380

GAAGATTCTG TTGTTTCATT TTCTGAAGTT GATTCAGCAG GTTTAGAATC TCTTGTATTG 1440

CTTGGTTTGT TTTCGTCGCT AGCAGTTTCA ATGTTAGATT CTGCAGTTGC GTTTGGTTGG 1500

TTCTCAGCAC TGGTGTTATC ACCATTTGCT TCAGCATTTC TTGCTGGACT TGTTTCTTCA 1560

CTTGCGCTAG CTTTTGACTG GATTTGATGA TTCAAAACTA GAATAGCTTT TGTCGATTCA 1620

AGTAAAGCTG TTTTGTCTTT ACTCTTAGCA GAAAGTTGAT CTAATAATGC ATCCACCTTA 1680

TCAAAGTCCG CATCAGATCC ATTATTACTT TCTAAATAAG AGTGAAGCGA CATGAGAATA 1740

TCGTAGAGTT TTTGATAGAG TACAAGTGTC TGAGGATCTT GCTCAGCATT TTCCTTTTCT 1800

TGTTGAAGGG CGCTAGCGAT ACGAGTCAAG ACATCTTTTA CCTGACTGTT TACTTCATCC 1860

AAGTCTGCAT CAGCCTTGTT TGTGGCAGCT TTTAGATTTT CTACTTCTTC TGCCAAGGAT 1920

TGTCTGATTC CTTCTTCATG GATTTGTTCC AAGAGTTGAT TTGCCTTGCT CAAAAGACTT 1980

TCTACTTCTT CCTTGCTATC TGTCGCAGAT TATTGGTTGC TATCTACCAT GTACTCCTAA 2040

AACAGGAGAG TTATAATCCA AGATTACAAG GCCTTACAGA AATAAGAAAT CCAGATAAGA 2100

CAATGTTCGT CCAAGACGCT ATTCGCTTCG CACAGCAGCA CGGATTCAAT ATGCTTTAAT 2160

TTTAAAGTTT AGGTGTCAAG ACCTCTTTTT AGTGTGCCCA AAATTTAGAG AAGTAATCAA 2220

TCAACTAACT TTTATTTTTT TCAAACTTTC AGTAAACTGA CCTAAAGCTA ACTCAATCTG 2280

TCTTTGTAGA TGCTTCTGCT ATCAGCTAGA AGTTGATCTA CTTTTGCCAA GACTGCCTTC 2340

TCATCAAAAG TTCCAGGTTG ATAGTTGGAT TGCAGGGATG GAATCTTGTT TTTCAAAGCC 2400

GCTTCATATC CCTTAGTTTG AACCTTGATG TAGTGATTGT GGTCGCCATG AGGAATCACA 2460

AAACCTTCTG AATCTTCACT TATAATTCGA TTGGCATCAA AACCATGACC ATCTTCTTCC 2520

TCATGATGGA CATGTAGTGA CGGATTACTT AATACAGAAC TAGAAGAACT TCCTACCTCT 2580

TCCGTGTTAG AGTGTGATGG GGGATTGTTA AGAGATGACT TAGGAATATA GTGATAGTGA 2640

TCCCCATGTC TTACTATATA AGCATCACCT GTATCTCTGA CAATATCATT AGGGTTAAAG 2700

ACATATGTGG CTGCTAATTC ACCTGCCGAC AAGTCACTCT CAGGAATGAA ATGATAGTGA 2760

CCACCATGTG GTACTATAGT AGATTGAAAT AGAATATGAG CAAATTGATA AGGGGATTTT 2820

AAAGTAATTT CTAACAATGA TTTAGAAACT ATGATGTGCT ATTCTAAATT CAACTCACTA 2880

TATATAACCA TCATCGGTAG TATAACGTCC CTGTAATTTT GCTACAGATA CTTCTGCACT 2940

AGCTCCTTTA TCGTCTTTAC CATGTTCTTG TTTTTGGCGA TTGATTTCAT CTTTTGTTCG 3000

TACATTTTCT GCATGAGCTT GATCTTTAAG GTAAACATAA TACTTTCCAT CTACCTTAAT 3060

AATATATCCT CCCTTAACCT AACTGACGAT ATCTTGATCT TTCGGCTGAT AGTTGGGGGC 3120

TTTCATTAAT AGCTCTTCAC TAAAGAGCGC ATCAAAAGGA ACTTTACCAT TATAGTAGTG 3180

ATAATGATCG CCATGAGAAG TTACATAACC TTGATCTGTA ATCTTAATAA CAATTTGTTT 3240

TGCTTGAATT CCTTCTTTTT GACTAACCTA GTCTGGAGTC AAATTTTCAG TCTTCTTAGT 3300

GTCTTTATTA CTGTTTACAT ATGAAACACG ATTTTTATCT GTATTGGCCT GTTAGCTATG 3360

TTGGTTCAGA GCATAAACAC ACAGACTTAA GGAAAGGATA ACAACAGATC CAGCTGCTAT 3420

ATATTTCTTT TTAAATTTCA TAATTACCTC ATTTCTATAA TTATTTATAT GATGTCTTCA 3480

TTATTAAATG ATTAAATAAA TTAATTAACC AATTAATTAA CTAGTAAATA TTCCACCTCT 3540

TTTTAAGTTG TATGTCAAGA AATTTTATAT ATTAATAATA AAATGAAATT CTCCCAAAGT 3600

CAGAGTTTTA TTTCTAACTT TTGAGAGAAC TTCATTTTTG ATTCAGACTT TTTCTACTGC 3660

TATTCCTTAC GCTATGAGAT CAGATAAATT CTTTTTTATC ACTTCTCCAC TTGGCAATCT 3720

TAATTCAATC GTTCCATCCA TATTGAATAT AACACTATCT AAGCCTAATC CGTAACTAGC 3780

TGTAAATTTT TCTAATTTTT CTTGTACAGG ATCTACTGCT GGAGCTTCCT CTAATGCTGG 3840

ATCTAACATA GGGTCACTCC CCACATTCCC TTCTGGATTC AACATTCCAT TATCCGTTGA 3900

GTTTTCTGGT TTTACAGGTT TTTCGTTTGG TGCCTCTGGT AAAGAATCTG CTGGTTTATT 3960

TTCTGTTGGT TGGTTCTCAA CTGTTCCAGT AGATACTTTT CCATTTTCAG ATGGTTTATT 4020

TTCACCATTT CCTTGAGGTG CTTCTCCTGT AAAATCTGCC ATATTCTTTT TAATGACTTC 4080

TCCCGATGGT AAATATAATT CAATTGTTCC GTCCATATTA AACAAGACAT TTTCTAGCTT 4140

CATCCCATAA CTTTCAGCAA ATTTTGCTAC TTTTTCTTGT ACAGGATCCA CTGTAGGAAC 4200

TTCTTCTAAC GTTGAATTAC TAGTACTATT CCCAGTTTCA GAAAGTTTTT CTTTTTCTAC 4260

CTTCTCACTA GTCTTTGGTT CTTCTACCTT TTCATCAAGT TTTAAGTTTT CTTGTGCTTT 4320

ATTCCTTTTA AATTGTGGTA GAATACTTGG TTTATCAGTT TGATTTTCTT TTTCCAAGAT 4380

AGGTACTTCC ACAATATAAG TCGATTGATT GTCCAAATAA GCATTTGCCA TGAAGGTTAC 4440

AGGAATTTTA TTTCCGGCCG TTCTGGTTGT TCCTTGGTTT AATTTCGGAA TCGGTAATTT 4500

GATTTCACCA ACTTTATAGT TATTTTCTAA ATAAGCATTT CCATGAAATT CATCAAACAC 4560

TCTGACTAAA GCATCAGTTC CTTTAGGCAC TGCAAATTGA GGGTTCACTC TTAAATAAGT 4620

ATCCCCTGCA TGGAAAGGAT AGAAAATCGT TTGACTGGCC ATTTTGTAAG CTAAAGAGGT 4680

1160

TGGAACTGTA AATGTACCAT CATAACTTAC TTCTGGATAA TCTTTTGAAG CGATAGTATA 4740

CTTAAATGTT TGTCCTGGTA AATAAGGTTG ATCTAATTCA AAGTTTGCAA TATTCCCTAC 4800

TCCTTCTCCA AATACTTTAC CAGATACTTT CTCCAATACT TTTCCATCTG GTGTTATTAA 4860

TTTTACTAGC ATATTGATAC CTAATTTTTT CTCCAATTCA GGCGGAAAAC TAAAAGAAAC 4920

GCGTTTTTGA CCATTGGCTA GAGTAAAGTT TTGATTATTA AACGTACTAT TTTTTAACAA 4980

ATTAACAACA TTCGTTAATT CTTCTCCAGT ATAAACTTTA TTCCCTTCTT TTTTAGCAAC 5040

TCCTTCTTCG GGTTTAAACA GTTCATAGTT ACTGTGAGAA TGACCAATTC CAACCGGTTT 5100

ATGTTCATCA ATCGGATCTG CATGATGGTG ATCTCCATGC GGATAAATAA TCGCATTTTT 5160

TTCTTTATTC ACGACAATAC TTTCACGTTT GACACCATAT TGTTTCATAA TGCCAGCAAT 5220

TTTTTCTTCG ATTTTTTTAT CTAAATCTTT CATTTCTTTG GCATTACTTG GATAATCCTG 5280

TTCATGAGAT GACAAAGAAT CTAATCCATT ATGACTAGTT TTAACTTCCT CTAAATGTTT 5340

TTGCGCAsCT TAATTTGCTC TTCTGTCAAG TCCTTCTTGA AGAAATAATG ATTGTGGTCT 5400

CCGTGACTCA TGACAAAACC TGATTCATCT TCAGCGATAA TACGATTAGC ATCAAATCCG 5460

TATCCATCTT CTTCATGTTT CTCATGTGAA GTTCCTGGAT TGATTGGAAG AGATGGAGAA 5520

GGTGTTGCTA GACTATTGTT TGGAAGAGTC GGTTGCCCAA TTTGATTTGA TTTTGGAATG 5580

TAATGGAAAT GATCACCATG TCTTACAATA TAAGCTGTAG CCGTTTCTTC AACGATATCT 5640

TTTGGATTAA AAATATAACC ATCAGATGCT GAAGAGAGCT CCTTACTTGT CGTTAAAGAA 5700

GAAGGATTGC TTGAAAGACT GCCTAGACTA GACACTACTT CATTAGGTTT TGCATTTGTA 5760

GAAACTGTAG AACCAGTTCC ACTGATAGGC ACCATTCTGG CAATCTTTTC TTCTAAGGCA 5820

GAAAGCTTGC TGTAAGGAAT AAAGTGGTAA TGGTCGCCAT GCGGAATCGC AACTCCATTT 5880

GGTGTACGAC TGATAATCTT AGCAGGGTCA AAGACCAGGC CATCTGATTC ACTGTAACGT 5940

TGGGCGCTAG GTGAATCATA GAGTTCCTTC AAAAGACTCT GGAGATTTTC AGATTTATTT 6000

GCTGGCTTGC TAGTTGATCC TTTTGCTACA GATTGCGTGT TATTGTCACT AGCTGTTGAA 6060

GAATAGCTTA ACTGACTCGG TTGCATATTT TTTCCAGCCA GATGTGCTTT AGCTGCTGCT 6120

AATTCACTAG CAGATAAATC GCTTTTGGGA ATGTAGTGAT AGTGACCTCC ATGAGGAACG 6180

ATATAAGCAT TACCCGTATC TTCGATAATA TCAGCTGGAT TAAAGACATA ACCATCATTT 6240

GTCGTATATC GTCCCTGAGA CCTTGCTACA GCAACATTAG AGTTAACCTT CTCATTATCT 6300

TTGACATGTT CTTGTTTTTG ACGATTGATT TCATCTTTAG TTCGAACATT ATCAGCATGA 6360

GCTGCATCTT TCAGGTAGAC ATAATATTTT CCATCGACCT TGATGATATA ACCACCCTTG 6420

ACTTCATTGA CAATATCAGC GTCTTTAAGT TGATAGTTTG GATCCTTCAT CAAGAGTTCT 6480

TCACTAAAGA GGGCATCATA AGGAACTTTC CCATTATAGT AATGATAGTG GTCACCGTGT 6540

GACGTTACAT AGCCCTGATC TGTAATTTTG ATTACAATTT GCTCAGCCTG AATTCCTTCT 6600

TTCTGGCTAA CCTGGTCTGG TGTCAAGTTT TCACTTTTCT GACTTGACTG GCTGCCATCC 6660

ACATAAGAGA CACGATTATT GTCCTTATTT TCCTGCGAAC GATGCTGGTT TAGTGCATAG 6720

GCACATAGAC TCAAGGATAC GATAACAGCT GATCCAGCTG CTATATATTT TTTACTAAAT 6780

TTCATAAATC CCTCATTTCA ATAAATGATG AAGTTTTTTC TCAACTTCTT TTACTTTATT 6840

AAATAGTTTT CTAAACCCGG GGGTACC 6867

(2) INFORMATION FOR SEQ ID NO: 193:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 999 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

CGTTCTAAAA ATGCAGTACG TTTGATTGAG AAATCAGTTA AAGGTATGCT TCCACACAAT 60

ACACTTGGAC GCGCTCAAGG TATGAAGTTG AAAGTATTTG TTGGAGCTGA GCACACTCAC 120

GCTGCACAAC AACCAGAAGT TCTTGACATT TCAGGACTTA TCTAAGGAAA GGAACAATAA 180

AGTATGTCAC AAGCACAATA TGCAGGTACT GGACGTCGTA AAAACGCTGT TGCACGCGTT 240

CGCCTTGTTC CAGGAACTGG TAAAATCACT GTTAACAAAA AAGATGTTGA AGAGTACATC 300

CCACACGCTG ACCTTCGTCT TGTCATCAAC CAACCATTCG CAGTTACTTC AACTGTAGGT 360

TCATACGACG TTTTCGTTAA CGTTATAGGT GGTGGATACG CTGGTCAATC AGGAGCTATC 420

CGTCACGGTA TCGCTCGTGC CCTTCTTCAA GTAGACCCAG ACTTCCGCGA TTCATTGAAA 480

CGCGCAGGAC TTCTTACACG TGACTCACGT AAAGTTGAAC GTAAGAAACC AGGTCTTAAG 540

AAAGCTCGTA AAGCATCACA ATTTAGTAAA CGTTAATTCG AAAGAATTAC TATACTTATA 600

CAGAGCACCT TTCGGGGTGT TCTTTTTTTA TACTTTCTTA CTAAATTGGT GCAATTGACA 660

CAGTTGTTGC GACTTTAGTC GCTTACAAAT GTGGCTGCAA CCTGACATGG TCAGTTGCCT 720

CAAAACGTTA ATCAATACGA TTATATCAAC GTTTCAAAGC ACTCAAGGGT TTACCCTATG 780

GGTGCTTTTT TCTATACTTT CTAAAAAAGT TTACCCTAAA ATTTGCCCTA AAATTACCCT 840

ACTTATTTTT AAGATGTTGG TAGGCAACTT GTCCAGCAGA TAATGGAACT ATGTTTGAAG 900

TATTAACATA AGTCTTAGTT GTAACGGTAT CGCTATGAGT TAATGCTTCA GAAATGGCTT 960

1162

CTAAGCTCAT TCCTGCTTTT TTAGCAAGTG TCGCTCCTG 999

(2) INFORMATION FOR SEQ ID NO: 194:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2315 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

AATATTATCA CTGTTCTTGA AGGCAGAACA CAAGCTGTCA TCCGAAATCA CTTTCTTCGC 60

TACGATAGAG CCGTTCGTTG TCAAGTGAAA ATCATTACGA TGGATATGTT TAGTCCTTAC 120

TATGACTTGG CTAAACAGCT TTTTCCGTGT GCTAAAATCG TTCTAGATCG TTTCCATATT 180

ATCCAACATC TCAGCCGTGC CATGAGTCGT TTTCGTGTTC AAATTATGAA TCAGTTTGAA 240

CGAAAATCTC ATGAATACAA GGCTATCAAG CGTTACTGGA AACTCATCCA ACAGGATAGT 300

CGTAAACTCA GCGATAAACG TTTTTATCGC CCTACTTTTC GCATGCACTT AACAAATAAA 360

GAAATTCTTG ACAAGATTTT AAGCTATTCA GAAGACTTGA AACACCACTA TCAGATCTAT 420

CAACTCTTAC TTTTTCACTT TCAGAACAAA GACCCTGAGA AATTTTTCGG ACTCATTGAG 480

GACAATCTGA AGCAGGTTCA TCCTCTTTTT CAGACTGTCT TTAAAACCTT TCTAAAGAAC 540

AAAGAGAAAA TCGTCAACGC CCTTCAACTA CCCTATTCAA ACGCCAAATT GGAAGCGACC 600

AATAATCTCA TCAAACTTAT CAAACGCAAT GCCTTTGGTT TTCGAAACTT TGAAAACTTC 660

AAAAAACGGA TTTTTATCGC TCTGAACATC AAAAAAGAAA GGACGAAATT TGTCCTTTCT 720

CAAGCTTAGC TTTTCTTCAA CCCACTACAG TTGACAAAGA GCCTATTTTC GCTGATTCTC 780

CACTACATTT GACTGGATTC TAATTTTTTA GAGAAATACA AAAGAGCTAG CTTTAGCTAG 840

CTCTTTTCCT ATGCGGAGAG AGGGACTTGA ACCCTCACGA CCTAAAGCGG TCACAGGATC 900

CTTAGTCCTG CGCGTCTGCC AATTCCGCCA TCCCCGCGTC GATTACTTTA CTAGTATATC 960

AACTTTTGGG ATGCTTGTCA ACACTTTTTT TCAAATTTTT TCATTTTCAC CAACCAGGTT 1020

ACTCAAAAAG TTCATTTAGA TTTTCATCTA CTAACTTAGC TCCGAGTGTA TTTTTGAAAT 1080

GACCTAGGGC AAATTGATGA TTTTCAGGCC AGATGGAAGC AACAGCTGGT TTAACAATCT 1140

CGATGTCATA TCCTAGATTA TAGGCATCTA TAGCTGTATG TAGGACACAG ATATCCGTCA 1200

AGACACCTGT TAAGATAACG GTAGACACTC TACGCTCTCT CAAACGAATA TCTAGGTCAG 1260

TCCCTGAAAA AGCTGAGTAA TGGCGTTTAT CCATCCAAAA GACACGACTG TCTGAACCAT 1320

GCTCTTGATA AAAGATCCCC AAATCTCCAT ATAAATTCCG TCCACTCGTC CCAATCAGAT 1380

TATGAGGAGG AAATAACTTA CTTTCCGGAT GGAAACAATC GTTTTCTTCA TGAGCATCAA 1440

TAGTAAAGAA GATATAATCT CCTCGTTCAA AAGCTAATCG AGTTACCTTG CTGATGGCAT 1500

CCGAAATCGC CTGAGCTGGA GCACCTGCTG TTAGTTTCCC ACTATCAGCA ACAAAATCTT 1560

CTGTATAATC AATCGAAATT AAAGCCTTTG TCATTAGTAA TCTCTTTTCT TCACTTCTTC 1620

AAAAATATCT GAAATCAAGA CCTTAAGATA GGTTCCCTTC ATTCCAAGTG AGCGACTTTC 1680

AATAATCCCC GCAGACTCAA GTTTACGAAG AGCATTGACA ATCACAGAGC GAGTGATTCC 1740

GATACGATCT GCAATCACTG ACGCAGTCAA CTTCCCTTCA TTTCCATTTA ATTCCCCTAA 1800

AATTGCTGAA ACAGCACGGA GTTCGGAGTA AGAAAGGGTA TTGACCGCCA TGGTGACAGC 1860

AGTACGACGA CGAATATTTT TCTCATCTTC TTCACGTTGG AAGTTAAGAA GCTGAATCCC 1920

AACAACGGTA CTGGCAATCT CAACAAGAAC CAAGTCCTCA TCTTCGAATT TTTTATCATT 1980

ACGCCAAATA ATCAAAGAAC CAAGGCGAAT CCCCGATACA TGAATCGGTG CAATAGTCGT 2040

CAAGCCATCT GGAAAATCAT CTCTACTCTC AATAGGGAAA ATACTCATAT CATGCTCAAC 2100

AGGCAAGTTT GCTTCTGTTT CGTAAATCAT ATTAGCCCCT TGAACGTAGT CATCTGGGAA 2160

AATCTTAGTT TGGAAGAATT GCTtACGCGA TCTGTATTTG TTTTATAACG CATAAAATAG 2220

CCAAGCAGAC GTCCCTTACT ATTGATAATG CAGGCATTGC AATGAATAAT ATCCGCTAAC 2280

TGACGCGTAA TAGCGTTGTA AGGGAGCTCA TCTCG 2315

(2) INFORMATION FOR SEQ ID NO: 195:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6693 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

CGATTTCTTC CATTTCTTCA AATAAGAATA CTTCATCTGA CATATGTGTT ACCTTCTTCA 60

TCAAAAATTA TTTTGTAATC GATTACATTG CAGATCGTAA CATAAAGAAA AACAGATGTC 120

AAATATTAAA CGTAAAAACA TGGTCACTAA AGAACTATAA GAGAAAAGGT AAACCTAGCG 180

ACGCGATGAA CGCTGGGTCG TTTGGTTTCG ATTGCTCTCT TCCTCTTGTT TTTTCTGTTC 240

TTCTTCTTGT TTTTTCTCAG CTTCCTTGGC CTCTTGTTTG GCTTTTTCCT CAGCTTCCAT 300

AATTAATTTA TCCGCCACAG TGTAGCTGTA GATTCCAGCT TCCATGTCGA CCACACTCGG 360

TTCTGACAAT TGAGGCTTAA TCTTACTGTA ATATGGCAGT TTCTTACTCA TTTCAGATAG 420

AGGAACCAAG ACTTCGTCCG AATCATTCAT GGTCAATCGA ATTAAATCGG ATGTCACCTT 480

GCTTGGGGCT AATTCCACCT TTTGGATAGC CGCCTTGAGT TCTGGGCTAA TTTGAGCAAG 540

TTCTGAGACA AAAACTTTGA TTTGTTCACT ATCATTAAAG AGAACTGATA AATAAGTTTC 600

TGGTAAACTG TTCAGACTCA CAGAACTAGT CTCAAGCTGA CCACTGGAAA GAATAGGATA 660

ATGATTTTCA CCAGAAATAT AGTAGGCCAC AATATCATAT TCCTTGACCT TAATAGTGAA 720

CTTAGTTGGA AATTGATAGA CAAGTTGAGC TGATTCAACC CAATAGTTAG ACTTAATCTG 780

CTTTTCATAT TTTGCCTTGT CTAGCAGAAG GTTAATCGTA TAATCCGAAT CCTGAATGCC 840

TGAAGCCTGT CGAATATCAT CAGCTGTAGT TTGCACCGTT CCCTCAACAC GAATATCTTT 900

CATGGTCGCA TAAGGACTGA GCAAGTAGGC AGAGACAAAC AATAGAAGCA GACTTGGAAA 960

TAAAATCGTG AAGGCTCGCA AGATATGGAT ACCAGGAATC TTTGCTTTGG CTGGTTTTTC 1020

CTTTGTAGCC TTTTTAGCAA GCTTTTTATC CTGTTCCTCC TTCTCTTTAG ACTCTGGTTC 1080

TTCTTTCTCT TCTTTCTCTT TGTCAGCCTC TGAGGATGCT ACTTTTTCTT CAGACTCTTC 1140

CTTAGCTGAT TCTGAATCTT CCTGGTCTGT TTCACTCTCC TGGTCCTGTT TATCCTCTGA 1200

CTTCTCAGAT TCTTCTCCCA TTCGAGCTTG TCTTTCCTTT TCCTTCTCCT CAGCTAGAGC 1260

CGCCTCTTCT TCAGCCTTCT TTTTTAGATA TTCTTGGTTT CGTTTCTGCC ATTCTGATAA 1320

CTCTTTCAAT TCTTCGAGGG TTTCTTTGTC CTCATTTTTC TTATCTTTTG ACATTTACTT 1380

TCCTTATGAT AAATCTTTTT TCAACAATTG ATAAAAATCT GCTAGAGATT TCAATTCCTT 1440

AGAAGCTTTC ATCTTAGCTT GGTAATCTTC CTTGTGACTT AGTAAGTGAG AAAGCTTCTC 1500

TTCCAAACTA TCCAAGGTCA AATCGCTTTC TTGAAGGTCT TCTGCATAGC CTTTCTTAAC 1560

AAAGTAAGCT GCATTTTCAA TCTGGTCACC ACGACTAGCT TCACGACCAA GCGGCACAAT 1620

GACATGCAAT TTTGCTATCG CCAAGAGCTC AAAAATCGTA TTGGCACCAC CTCGTGTCAC 1680

AACAATATCA GCCAATTCCA TCAAGGGTTG ATAGAGATCG GTCACATAGT CAACACGAAA 1740

AAGATTTTGC CTCAACTCAT TCAGACTAGA ATCTCCAGTT AGATTGATAA TATTGTAGCG 1800

CTCTGTTAGT TCTTTCTTAT GGTCTGTCAC CAATTGGTTA AAGACACGAG CGCCTGCAGA 1860

ACCGCCAACA AACAATACAG TTGGCAATTT GGGATTAAAG TGGGTTTGAA TATCCACCAA 1920

TTCATCTGGT TCTGGAGTGT TTTTGTCCGA AACCTTGGTC ACCGCTCCCA CATGCTCAAC 1980

CTTAGCCAAA CTCGAAGCTT GTTCAAAGGT TGAATACATC TTAGTCGCAA ATTTATAGGC 2040

GATTTTATTG GCCAAGCCCA TAGACAGGTC AGATTCGTGA ATAAAGACAG GCACTCCTGA 2100

CACACGCGCA GCGATAACAG GCGGTACTGA GACAAAGCCC CCCTTTGAAA AAAGGGTCTG 2160

TGGACGCAGT CGCAACATGA TAAAGAGCGA TTGGACAATT CCCCAACCAA CTTTGAAGAC 2220

GTCCAGCATA TTTTGCCAAG AGAAATAGCG ACGCAATTTT CCAGTCGCAA TAGAATGGAA 2280

GGTGACATCC AAACCTGACT TAAGGATTTC TTGGTGTTCG ATACCACACT TGTCCCCGAT 2340

ATAGTGGACT TCCCAACCAT CTTCGATGAA CTTGGGCATT AACAAAAGAT TGAGGGTAAC 2400

GTGTCCAACC GTCCCCCCAC CTGTAAAGAC AATTTTTTTC ATATTATTCT TTTAACTCCG 2460

CTACTGTGTC GATAAAGAGG TCGCCACGTA CTTCAAAGTT AGCATACATA TCCCAGCTAG 2520

CATTGGCAGG ACTAAGAAGA ACCACATCTC CTTGAGTCGC AAGCTCATAG GCCTTGCGGG 2580

TCGCATCTGC AATATCTGTC GCCTCCACAT AAGCGACACC AGCCTTGTCT GCTGCCCGTT 2640

TGACACGTTC TGCAGATTGA CCCAGGATGA CCATCTTCTT GAGTCCAGTA ATGTCTGGCA 2700

CCAATTCGTC AAACTCATTG CCACGGTCCA AACCACCTGC AATCAAGACG ACCTTGCTGT 2760

TGTCAAATCC TGACAAGGCT TTTTGAGTAG CCAAGATATT AGTTGATTTA CTGTCGTTAT 2820

AGAATTTAAC ACCCTTGATG TCATCCACAA ACTGGAGACG GTGTTTGACA CCACCGAAGG 2880

CTGAAAGAGT TTCCTTGATG GTTTGATTGT CCACATCACG AAGCTTGGCT ACAGCAATAG 2940

TCGCAAGGGC ATTTTCCACA TTGTGGCTAC CTGGAACACC GATTTCATTC GCTGCCATGA 3000

CTACTTCACC ACGGAAGTAG AGTTGACCAT CTTCCAGATA AGCTCCATCA ACCTTTTCAA 3060

GTGTTGAAAA TGGTACAACA GTGGCTTCTG TCTTGGAAGT CAAGTCTTTT GCCAAGTCTT 3120

GATTAAAGTT CAAGACAAGG AAATCAGCTG CTGTCATCTT GTTCTGGATA TTCCACTTGG 3180

CTGCTACATA TTCCGAAAAT GACCCATGGT AGTCGATATG AGTTGGCATG AGGTTGGTAA 3240

TAACCGCAAT CTCTGGATGG AATTCTTGAA CACCCATGAG TTGGAAAGAA GAAAGTTCCA 3300

TAACAAGCGT GTCCTTATCT GATGCTATTT GAGCAACCTG ACTAGCTGGA TAGCCGATAT 3360

TCCCTGATAA AAGACCATGT TGGCCAGCAG CAGTCAAAAC TTCCCCAATC ATAGTCGTTG 3420

TGGTTGTCTT ACCGTTCGAT CCTGTGATAC CAATAATCGG TGCTTCTGAA ATCAAATAAG 3480

CCAATTCCAC CTCAGTCAAG ACTGGAATTC CCTTGGCCAA AGCCTTTTCA ATCATGGGAT 3540

TGTTGTAGGG GATACCTGGA TTTTTCACCA TAAGGGCAAA CTCTTCATCC AAGAGTTCCA 3600

AAGGATGGCC ACCTGTAATG ACCTTGATCC CTTCTTCCAG CAAACTTTGG GCAGCTGGAT 3660

TGTCCTCGAA AGGTTTCCCA TCATTTACTG TCACAATGGC ACCTAGCTTG TCCAACAAAC 3720

GAGCTGCAGA TTCACCAGAC TTGGCCAAAC CTAAAACAAG GACTTTCTTA TTTTTAAATT 3780

GATCTATTAC TTTCATGTCT CGAACTCCAT TTCTACTCCT ACTATTTTAC CATTTTTATG 3840

GAAATAAAAA AGCCACAAAG TGTGTTTGTG ACTCTTTCTT CTAACTGAAT CTTACCATAT 3900

CATCTATGTG ATAAATCGGT AACTCGAATG ACCTGATCCA CTTGCTCCCA AATCAGAGGA 3960

TTATGGGTCG CAATAATAAT GGTCCGATTC GGATTTTTTA AAGATTCTAG GATGGAAAGT 4020

AATTCCTCAG AGTTTTTGGG GTCTAAGGAA GCGGTTGGTT CATCTGCGAG GATCAAAGGT 4080

GGATCCTTTA AAATTATCTT CGCTAGTGCA ACACGTTGTG CTTCTCCTCC TGATAACTCA 4140

AATATAGGTT GCTTCAAATC CAAATAAGAG AGGTTTACAC GGTTTAGAGC TTGTTTCATC 4200

AAAGAGATTT TCTCTTTTTC CTTCAACTTT TTACCAACTA AACCCAGATT GAGATTCTCT 4260

TTGACGGTTT GGCTTTCAAT TAAGCCAAAA TCTTGAAATA AGTATCCTAA GTAATCTCTA 4320

AAGAAAACAG AAGGCTTGAT GTCCTTAAGA GAAGTGCCAT CATAGATGAT TTGCCCTTTG 4380

TCATATGGCT CCAATCGTCC AATCATATTC AAGAGTGTTG TCTTACCACA GCCACTTGTA 4440

CCGATTAAGG CATAAATTTT CCCACCTTCA AAATGAAGAT TCATATCTGA AAATAGCTGA 4500

CGGCTTCCAA ATTTTTTAGA TATATTCTTT AGTTCAATCA TCCTATTTTC CTTTCATAAT 4560

TGTCATAGAA ACACGAGATT CTTTCTGCGC TTGACGGTAA AGCGTCAAAA CTGCACTAGC 4620

TAGAAAGACC AATAAAGTGA GCAAGCCAAT CACCAAGTCT CGACTGCTTA AAATAAAGAG 4680

ACTAGCACCA AATACAAAAC TAGCAAATTG GCTAACCATA TACTGAGCAT GTGTTTCAAA 4740

AAATCGTAAA CCTGAAATTC GTTTAATCAA GATATCTCGG CGGAATTGCT CGAAATATAG 4800

AAGATTGACA GAATAAAAGA GTAACAAGGA ACTGGCTATT CCAACAATAG CTCCTAAGAT 4860

TAAAGTTGCT GTTTCAGTTT GAACTTCATT ATAACGAGTT AGATAAACAC TTCTTCCTTC 4920

TTTAAGATAG GATACTTGCT CATAAATTCC AGCTTTCTTC AAGAGTTCTA GCCCACTCTC 4980

ATATCCTTTG ATAAAGAGTT GTTTTCCAGC ATTGATAGAC CAACTAGATA AGGATATAAA 5040

ACTATCACCT GTAGAAGTCG GCGTGAATAC CACTAAAATC GGATCAGTCA AATACTGAGT 5100

AGATACGGGA TTCTCACCGT TATTATAAAC AAACCGCTTT TCTCCCATTG AAAGATAACT 5160

AACGTGCGCT TTCATCTCAT AATCCAAAGG AGCACTTGCC TCCTCACCAG ATTTTCCATA 5220

ATAACTCAAT CTTTCTTCAA AAACTTTCTT AAGTTCTGCT TCTCGAGAGC GCAAATGTTC 5280

TGGGAGCAAG AGGATAAACT CACCTTTTTG GAGATGGGCT AACTTCTGTT TGGTCTCAGC 5340

ATCTACCACG ACCTTTTCCT TGTCCAAATA ACTGGGACTA ACATAGAGCG TATTAGCATC 5400

TGAACTATAG GTATCCAGTG TCTCTCCCTG TTCATTTTTT CCTTGTGGAT TGGCAAAATG 5460

GAGCAGATTA TCCTTTACAT AAAGAGCTTG TTCTTCTTCG ATTGCTTCCT TGGCAAAGGC 5520

ATACCACTTG CTCTGATTTT CTGTATCTTT TCCTCTATCA CCTAAGCCAA AGGAAATCTG 5580

GTAATAGTCT GCTCTGTCCT GCCATGCTTG TTTTGAAATT TCAAGTTCTT TCAATCGTTG 5640

GTAAGACGTC AAACCTGTCT TAACAGCGTA GCCTACTGTA AAAACAGCTA CTAACTGACA 5700

CAATAGGGTT AAAGCCATCA AGCGTTTAAG GGGTAATCTT CCCTTAATAA CGGGAACTAA 5760

TGCTTTGTAA CTCAAACTCA TTAGGTAAAG GAGCATTAGT AAAATTGAAA TCGCCAATAA 5820

AAACAACAGA TAGAAACTAA TCCCAAAACC ATAGGTGGCT AACAAGATAG GATAAAACAA 5880

ACCTTGACTA AAAAGAACGA CTCCCCCACC TAGGAAGGAA AGGAGGGCTG ATAGAAGGAG 5940

CCATTTGATA TCAGTAGATA AAGAATGCCC CATGATGGAT AAGAGAGTCT GACCAGAAAA 6000

GAGTTTTATA CCTGCTGCTC TCATTTCCTT AATCCGAGTG ATAATCACTA AAGCAAAGAA 6060

AGATAAGCCA AATATTGCTA AACTAATTAA AATAAGGGGA TTTAGTAATA TTCGAAAAGC 6120

AAGAAAATAG GGCGGTATCT TTCGGTCAGC ACTTGCTTTA TAACCCAAAT CTCCTAATTT 6180

ATCGGCAAGC TTTTCTTTCG TCAAGGAGCC TGACAAAAGG AGATAACTAT TTAGCGGAnT 6240

AtACGTTCAC GACTTTCTTG GCTAGCTTCT TGGAATTCTT TTGGTAAAGT TCCCTGACCA 6300

TAAGTTGCAT AAGTAAAGTG AGTCGTCCCA TCCTTACTCG GCTCTACAAT TCTTCTAGCT 6360

ATTAAACTCT GTTCTGAGTT TGCAAAATTC TCCAATTCCT GTTCAAATAC CTCACGCGTC 6420

GGTTCCTGAG TATCTTTTTT GACACGAAGT AAAGAAACGG AATCATAGCT TGCATATAAA 6480

TATTGTGGCG CACGTAAGAC AATAATCCAA GCAAGGAAGA AGCTGAGAAA AAAAGTTGAT 6540

AATAATATGA ATAGTTTCTT CATAGTAGAC TCCTTGTAAA CAAAATTCCC CCTGTAATTT 6600

CTTACAAGGG GAACGATTTA AATCAATGAA CGATTAGTCA TAATCACAGT AAAATGCTAC 6660

TTGTTCTCCC CATTTAGTCC AAATCCATGC AGG 6693

(2) INFORMATION FOR SEQ ID NO: 196:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1847 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 196:

CCGGTCTATG TACCCACTAC TTTGGGACAA TATGGGGATC AGCTACCCAA AACTAATCGA 60

GCGTTTGGTT GACCTTGCCA AGGAAAGTTT TGACAAGCGC GACGATTTGA TATAAAATGA 120

AAGAGAGGGT AGAAGCCAGA ACCATCACTG CACGGTGACT AGAGTTCTCG GACTTCAGCC 180

CTTTTTAAAG GAGTAGAAAT GAAATTAACA ATCCATGAAA TTGCCCAAGT TGTTGGAGCC 240

AAAAATGATA TCAGTATCTT TGAGGACACC CAGTTAGAAA AAGCTGAGTT TGATAGTCGT 300

TTGATTGGAA CTGGAGATTT ATTTGTGCCA CTTAAAGGTG CGCGTGATGG CCATGACTTT 360

ATTGAAACAG CCTTTGAAAA TGGTGCAGCA GTAACCTTGT CTGAGAAAGA GGTCTCAAAT 420

CATCCTTACA TTCTAGTAGA TGATGTTTTG ACAGCCTTTC AATCCTTAGC ATCCTACTAT 480

CTTGAAAAAA CGACTGTTGA TGTCTTTGCT GTTACAGGTT CAAATGGCAA GACAACGACT 540

AAGGATATGT TGGCGCATTT ACTGTCAACA AGATACAAGA CCTACAAAAC ACAAGGCAAT 600

TACAATAATG AGATTGGCCT TCCTTACACA GTTCTTCATA TGCCTGAAGG AACAGAAAAG 660

TTGGTTTTGG AGATGGGACA GGATCACTTG GGCGATATTC ATCTCTTGTC TGAATTGGCT 720

CGTCCAAAAA CAGCCATCGT GACCTTGGTT GGAGAAGCCC ATTTGGCCTT TTTCAAAGAC 780

CGTTCAGAGA TTGCTAAGGG AAAAATGCAA ATTGCAGACG GAATGGCTTC AGGTTCCTTG 840

CTTTTAGCGC CGGCTGACCC TATCGTAGAG GACTATTTGC CAACTGATAA AAAGGTGGTT 900

CGTTTTGGGC AAGGGGCAGA GCTGGAAATT ACTGACTTGG TTGAGCGCAA AGATAGTCTG 960

ACCTTCAAGG CCAATTTCTT AGAGCAAGCC CTTGATTTGC CAGTAACTGG CAAGTACAAT 1020

GCGACAAATG CTATGATTGC ATCCTATGTT GCCTTGCAAG AAGGAGTTTC AGAGGAGCAA 1080

ATTCGTTTGG CCTTCCAAGA TCTTGAATTG ACGCGTAACC GTACCGAGTG GAAGAAAGCA 1140

GCCAATGGAG CAGATATCCT GTCAGATGTT TACAATGCCA ATCCAACTGC TATGAAACTG 1200

ATTTTAGAGA CTTTCTCTGC CATTCCAGCC AATGAAGGTG GCAAGAAAAT TGCAGTGTTG 1260

GCGGATATGA AGGAGCTTGG TGACCAGTCT GTTCAACTTC ATAATCAGAT GATTTTGAGC 1320

CTTTCTCCAG ATGTGCTTGA TACCGTGATT TTCTATGGAG AAAATATTGC TGAATTAGCC 1380

CAATTGGCCA GTCAAATGTT CCCAATCGGC CACGTTTACT ACTTCAAGAA AACAGAAGAC 1440

CAGGATCAAT TTGAAGACCT AGTCAAGCAG GTCAAGGAAA GCCTTGGAGC CCATGACCAA 1500

ATCCTGCTCA AAGGCTCTAA CTCTATGAAT CTAGCCAAGT TGGTAGAAAG TTTAGAAAAT 1560

GAAGACAAGT GATTTTGTCA AGTATTTGCA AAGAATGATT GCCATTACAG ATACTGGCTT 1620

AACCTTTACA AAAGATCCGT TTGACCGTGA GCGCTACGAA GACTTGCGAA GTCTGTTATC 1680

TGAAATGTTG AATCAAGCAT CAGACCTTGA TTCCGAAGAA GTGGCAGAAG TCTTGAAGCC 1740

AACTTCTGCT TATGCGACTC CGTTAATGGA CGTCCGTGCT TGGATTGTTG AGGATGAGAA 1800

GATTTGTCTG GTTAGGGGAC AAGGAGAGGA TAGTTGGGCT TTGCCGG 1847

(2) INFORMATION FOR SEQ ID NO: 197:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1062 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

CAAGCGAAAA CATTTTTTAT TCCAAATAAA CAGAGCATTT TAGGAGAACA AGAGATTTTG 60

AATGCCAAGT CGATCTTGGC CTTGCTAGAC GGTTTGGAGT CACATAGCTA TGATGTAGTC 120

TATCTCCGTC AGCCTCTTAA TCGTCTCGAA TATATCGAGT GTGCGATAGT GGGGCAATCA 180

CAATTTCTCT TTAAGGTCAG TTATGCTGAT GGTCAAAAGG CTTACCGTGT CGATCTTCCT 240

GACCTACTAA CAAAGACAGA CTGGCAGATT ATCAAGTCAT TTTTAGATGC TTTGCTTGCT 300

TATACAGGGA CTGATATTGA AGGGCTAGAT GGTTTTGATT TTGAAGCTTA TTTCCAAGCA 360

AGTATTCAAG CCTATCTAGC AGACCCTGTA GCTCGTTTTA CGATTTGCCA AGGAATTTTT 420

AATCCTATTT TCTTTAGTCG TGAGAACTTG AAAAGCTTTT TAGAGGCAGA TGGCTTGGCT 480

CAGTTTGAAG CGCGTGTGCG TGCGGTTCAA GAGACAGATG CCTACTTTGC GAGAGTTTCC 540

TTCTATCAGG ATGGAGAAGG AAAAGTGCAT GGCGTTTACC ATCTAGCTCA AGGAGTCAAG 600

ACAGTTTTAC CGAGAGAACC GTTTGTTCCT GCAGCCTATA TTGAGCAATT GGTGGATAAG 660

GAAGTCCAGT GGGAGATTGA CTTGGTTCAA ATCACAGGAG ATGGCTCTAA ACCAGAAGAC 720

TATGAAGCCA TTGCTCGCTT GGACTATGCA AAATTCTTAG AGGTATTACC CCCATCTTTT 780

TACCACCAAC TAGACGCCAA TCAAATAGAA GTGCAACCCA TATTAGACAA AGATTTTAAA 840

ACATTAGCAC AAGAAAAGTA AAGCAGAAGC AGGTCAATCG ACTTGCTTTT TTGACATAGA 900

AAAAATCCTG CCAAGaTGAC AGGATTGCTA CTCAATGAAA ATCAAAGAGC AAACTAGGAA 960

GCTAGCCGCA GCTGTACTTG AGTACGGTAA GGCGAAGCTG ACGTGGTTTG AATTTGATTT 1020

TTGAAGAGTA TGAAGTTTAA AGAAAAGCCA AGATACGAAG AT 1062

(2) INFORMATION FOR SEQ ID NO: 198:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6846 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

TATCTACAAC CTCAAAAACA TGTTTTGawG gCTCGTCAGT cTATCTACAA CCTCAAAAAC 60

ATGTTTTgAa kGCtcGTCAG tTCTATCTAC AACCTCAAAA ACATGTTTTG AcaGCcTcGT 120

CAGTTCTATC TACAACCTCA AAAACATGTT TTGAGCTGAC TTCGTTAGTT TCATCTACAA 180

CCTCAAAAAC ATGTTTTGAG CTGACTTCGT TAGTTTCATC TACAACCTCA AAAACATGTT 240

TTGangnCnT CGTCAGTTCT ATCTGCAACC TCAAAGCAGT GCTTTgagcG CTTCGTCAGT 300

TCTATCTACA ACCTCAAAAC AGTGTGTTGC GCAGCCTTTA ATCAGCCGCC TAGTCCGCTC 360

TATGGTATTC ATTAAGTCAA CATCTCTTGT TTAAGAGCAC CAAATCAGGA AATCTTCTCG 420

ATTCCCTGAT TTTTTCTATT TACGTTTTCG TGTTGAGCTA CGTTCTGTCA AACCATGAGG 480

TAAGAGAACT TCACGTTCTT CCAACTCTTC CTTATGCATA ATCTTGGTCA ACATACGCAT 540

ACTAATGGCA CCAAGGTCAT AAAGAGGTTG GGCAATCGTT GTCAAGTTTG GACGGGTAAA 600

GCGTGAGATT TGTGAATCAT CACTAGTAAT AATTTCAAAA TCTTCTGGCA CAGAAACACC 660

CTTATCAGCC AAACCGTTCA AGACTCCTGC TGCCAACTCA TCACCTGTCA CAACTGCTGC 720

AGTTGCATTT GATGAAATCA AACGCTCTGC TAAGGCGTAA CCATCATCAT AGCTATATTT 780

AGATTCAAAT ACCAAACCCT CACTATAAGT GATTCCTGCT TTTTTCAAGG TTTCCTTGTA 840

GCCAACTAAA CGAACCTTAC CATTGATGTC ATCCACTAGC GGACCGCTAA CGAAAGCAAT 900

ACGCTCATTT TCTTTAGCAA GGTAACTCAC TGCATCAATT GTTGCTTGCT TATAGTCAAT 960

ATTGACACTT GGCAACTGGT GCTCAACATC GACAGTTCCT GCGAGAACAA TCGGAGTACG 1020

TGAACGCGAA AATTCTGAGC GAATTTTATC TGTCAAGTGA TACCCCATAT AGATAATGCC 1080

ATCTACCTGC TTTGAAAAGA GGGTATTGAC AACAGAAACT TCTTTCTCGT TATCTTCATC 1140

GCTATTAGCT AGGACAATAT TGTACTTGTA CATTTCTGCA ATATCATCAA TCCCCTTAGC 1200

CAAACTCGAA AAATAACCAT TGGTAATATT TGGAATCACG ACACCGACAG TGGTTGTCTT 1260

TTTACTTGCA AGACCACGCG CAACTGCATT TGGACGATAA TCCAAACGAT CAATTACCTC 1320

TAGCACTTTT TTACGGGTAT TCTCTTTTAC ATTTTTATTG CCATTGACCA CACGGCTGAC 1380

CGTCGCCATG GAAACACCTG CTTCACGAGC GACATCATAA ATGGTTACTG TATCATCTGC 1440

ATTCATTCCT TTTCCTGTCC TTTCTATCTC ACACATTCTT TTACAAGTAG AGGTACTGAT 1500

TGAAGCTCTA TATCTACTTA CAAAAGTGAA GATGTGAAAA TTTCGTTTTC ATATTTCTAC 1560

TTATTCCATT CTATCACTAA TTGTAAACAC TTTCAAGTGT TTTTTGAAGA TTGATTGAAA 1620

AAATTTCATA GAAAACCTAG GTTTAGCTCC TTGCTACCAC CTTAGACTAA ACAAAAAGGA 1680

GGAAACTAAG CCCTCCTAAA GTTATAGTAA AATGAAATAA GAACAGGATA AATCGATCAG 1740

GACAGTCAAA TCGATTTCTA ACAATGTTTT AGAAGTAGAG GTGTACTATT CTAGTTTCAA 1800

TCTACTATAG GTATTGTTCC ATTCACTACC GTCAATTTTA GCACATAGTC TTCATGAAAA 1860

TATTATATCA TCATAACCAA CCAGATTCTT TCGCGATATT AGCTGCCTCT GTTCGATTAC 1920

CTGCATCTAG TTTCGAAAGA ATATTGGTGA CATAGTTTCG GACTGTTCCG TTGGATAGAT 1980

AAAGTTTGTC TGCAATTTCT TGGTTAGAGA AGCCCTGAGC AATTCCCTTT AAAACTGCGA 2040

TTTCTTGCTC CGTTAATGGA TTGGGATGCA TCATCACCAC TTCCATCAAT TCAGGCGAAT 2100

ACTCCTTGCG TCCTTCGAGG ACGGTGTGCA AGGTTTGCAT GAGGTCTGCA ATGTTTCTTT 2160

CTTTTAATAC ATAAGCATCT ACTCCAGCCT TGACCGCACG TTCAAAATAC CCAGGACGCT 2220

TGAAGGTCGT CACCACAACC ACCTTTGTTT CAAGCTTTTC TGCTCGTATC CACTCCAAGA 2280

CTTCAAGACC TGTCTTAACA GGCATTTCTA CGTCAAGGAT GGCGATATCT ACAGACTCCT 2340

TTTCTAATAG TTGGATTGCT TCTTGCCCAT TCTTGGCTTG AAAGACAGAC TCTACATCCG 2400

GTTGAAGCAT GAGCAACTGG CACATGGCAT CTCGCAACAT ACTTTGATCT TCTGCGACTA 2460

ATACTTTCAT CTACTTTCTC TCCTTATAAA GTAGTCGAAC CTGCACTTCA GTTGGATGTT 2520

TCTGACTGAT TACACTTACT TCTCCTGAAA ATGGAAAAAC ACGATTTCGG ACTGTATGGA 2580

GCTCATCCCC GCTTATAGAG GCAAAGCCAC AGCCATCATC TCTCACTGTT AGAATGAGTT 2640

CTTTCTCTGT CCGTTCTAAT TTCAAGTAGA CTTTAGACGC TTTAGCATGT TTGATGATAT 2700

TGGTCACTAA TTCAAGCAAA ATCATGGAAG CCGTTGACTC CAATTCCTGA GTTAAGCTAG 2760

ACTTGTCCAA GTGATTCTCA ACTTGAACCT CAATTCCAGC AATTTCTAAC ATCTTTTTCA 2820

CAGTCTCTAG TTCGGATGTC AAAGTTCTAG ACTTAAGATT TTCCACAATG GTTCGCACTT 2880

CATTCATGGA tCCTTGCTGA TCTGGTGAAT TTCTTTTAAT TCCTTTTCCA CCTGTGGATA 2940

AGCCTCCATC TGAAATAACT GCAAGGCTAA ATCTGTCTTG ACACTCAGCA TAGCAAAGGT 3000

ATGTCCCAGA CTATCATGCA AATCCTGACC GATACGACTA CGTTCATTTT CAGCAAGCAA 3060

TAGATTTATC TGAGCATTTT GCTTGACCTG AGCTTCTTTC AAATCCTCGA CAATACGAAT 3120

CCGAACCAAT CCAAAAGTCA TTAAATCGAC AAAAGTAAGA ATTACAAGTA GATAGAATAG 3180

AAACTCAACT TCGATTCTCT GAAAAATCAA CAGTTGCCCC ACAACAAGGA CTTGAGCAAG 3240

AAGAAAAGTC CAGACATGTA AAGACTTTAA ACTACGTACG CTGAAATGAT AACTTAAGAG 3300

ATTGGATAGG AAAAAGAAAA ACCAGATATA ATTAACAGCA ACAAAGGCAG TATTCCCAAC 3360

TACATAAGTC AGCATGAGGC CCCAATATAG CCAAGATAGG CGCTGGCTCT TAGTTGTTAA 3420

AACACCCAAA TATGCCACTA CAAATAGAAT ATCAATCAAT AAATGCCAGG CAGAAAGCCA 3480

CCCAGTCACT ACAGACAGGA TGGGGAAAAT CATAAAAATT AAACTGATCC AAAACATATA 3540

ATGTATTCTT TTCAGTCTTT CAAGCATTAA GCATTCTCCT TATGACCTTG AAGGTAAATG 3600

GTCAAACCAA ACAAAACTAC TGAAAAAACA AGTAAATAAA CTGTGGCTGA TAGATTGATG 3660

CCACCCTCAT TTAAGAAGGT CTTGAGCAAC TCCATCAACT GATAGGTCGG GAGACACTTA 3720

CCTACTACTT GCATCCAGTC TGGAAATAAA GAGATAGGCA TCCAGAGTCC ACCTAAAACA 3780

GCCAACCCTA GATAAAGAAG ATTGCCCACG ACAGACATCA ACTGACTAGT TGGTAAGAGA 3840

GTCAAGGTCA AACCAAGCGC TACGAAGGCA ATACTTCCTA CTATCAGCAA AAGTGCAGCC 3900

CCAATCCAAT TTCCAAGAGA CATGTCCACA CCTCTTACAA AATGCCCAAC TGAGAAAACC 3960

ACCAAGATTG AAACCAAATA ATCAACCAGC ATACTTGTTA TCTTTGATAG ATAATATTCT 4020

ACCATATTTA CAGGGCTATG ACGCAATGTT TTCTGCCAGT TGTTGATCTT GTCGGTATGT 4080

AAAACAACTG GGAATGAGAA GATAGCTGTT GACATCATGG AAAATGCAGT CATGGAGATA 4140

AGATAATCAC GCATAAAATT CGCGAGTTCA CCTGGTGTGT CCTGATAGAT ACCAGAAAAA 4200

AATAAATAGA AAGCCGTCGG CATCCCTACT GACAATAGAT AATAGATCAA TTGTCGTTTG 4260

GTCAATAAAA ATTCTATCTT ACTAAGTGCT AGCCATCGTT TCATCTTAGT TATCTCCCTT 4320

CTGCGTTTCT TCAAAGATTG TATCCAACAA ACTACGATTA TTAACTTCAA TTTCTTGTAT 4380

GCCACATCCT GCTTGAACTA ACAGTTCCCA AAAAGCATCT GCTTCGCGTG TGACTACTTG 4440

TAGAGCATCC TGTTTTTGTG ACCAGTTTTC AACCAAGTTA GACTGCTCAA TGACTTCCTT 4500

GTATGCCAGA GGAAGGATAA AATGCTTTTC AATTCCCTCA CTACGCATAG CTAGAGGCGT 4560

CGTATCACGA ATCAACTCTC CCTTATTTAA AACCAAAATC CGGTCAGCCG TATGCTCTAC 4620

CTCTTCAATA TAATGAGACG AATAGAGAAT CGTGACTCCT TGCGCTTTTA GGTCCCGAAC 4680

GATTTCCCAA AAGCGTTGAC GAGTTGAAGT ATCCATGGCA GCAGTTGGTT CATCTAAAAA 4740

GACAAGCTTT GGTCGCCCAA TCAAGGTCAA GACAAAAGAG AAGAGACGCT TTTGCCCGCC 4800

TGACAATTTT TCTGCGAATT GCTCTTTTTG TTGCTGGTCA AACTGCAATA GTTGATCGAT 4860

TTCCTGATCG CTCAAGGAAT TTGGATAGAT ACGTTGAAAG AAAGCAATCA ACTCTTTGAC 4920

CTTTAATTTC TGAACGATGA CATTTTCTTG AGGCAGATAA CCTCTAATAT AGTCTAACTG 4980

AGAACTCGTC ACTGACAAGC CTTGGATGGA TACTTGACCG CTTGTGACCA GTTTATCTCC 5040

AAGCAGACAG TCCAAGAGTG TGGTCTTCCC AGCACCATTG GGCCCAATCA AGGCGACGCA 5100

TTCACCTTCA GCTACCTCAA AGGAAATACC CTTCAAAATA GCCTTGCCCT TGATGTTTTT 5160

ATTTAGGCTT TCTACCTTAA TCATATTCAT GATATTCTCC TTTCAACCAC TCCATTCTCA 5220

TAAGGAAAAC GACGAAAATC ATAAATCCAA ACCCCAAAGC ACCACGAATG AATTGGCGAA 5280

gCAAGGTTTG GTCAAACCAA CCTGTAAACA TTTCCACTAA CCATACCAAG AGTGACAGGC 5340

CGATAAAGAA ATAGATGATC CCTCTCTTCA TTCCTCAAGC TCCTTTTTCA CATCTCCGAC 5400

TAATTTCAAA CCTTCTCTAA CAAGCCAAGA CATCATTCCA AAGCCAGCAA AGAGCTCCCA 5460

AGGAAAATGA TAGAAACTCT CATCCAATCC CGAAAACATG AGTTAGGTCA TAACTCCTGC 5520

TACTACTAAA CTCACTGCGA TAATCATTTT ATTTCTCATC TCTTCTTCCT CCATTTCATA 5580

CTACAATTAT AGTCTTTTGA AATCAGAGGA GACAGAAGCT TCTGTCACTA GAAAATATGA 5640

CAAATGTCAT AAAAAATTCT GTTCAAAACA AGCAAGATAC ACTATACAAT AAAACACAAT 5700

TAGAAAAATC TAAGGCAACT TCCTCAAAAG AGATATCAAA CCCAATTCAC ACCATAATGT 5760

AAACTAATAC TTATTTAAAA TCAAAAAGAG TAGAAATTTT TATCAGACAA ACACATATAT 5820

AGTGTATTGA ATCTATAACA GTAGGCCTTA AATACTAAAA TATTTCTATA AATTAATTTA 5880

ACTTTCCTGA TAGAGCTGTT CATATCTTAT TTCAATTCTC TAAATTATAC GTTGAACAAA 5940

ACCCTTCTAT TTCTTTCTTA AAGATTTATA AGAGTTATAA AATCTGTTAA ATTTCAATGT 6000

GTATACCTAA ACTACGGTAT TTATTGAAAA GACTGGAGAC AAAAAGTATA CGCTGCCAAA 6060

ATGAATTACT GAAAATCAAA AAAGAGAGAA CCAAACTGAT TCCCTCTTAA TGTATATAAT 6120

ATCTAGTTTT AAAAATACAC ACTCACATAT CTCTGTAATG AATCGGGAAG ACAGGATTCG 6180

AACCTGCGAC ACCTTGGTCC CAAACCAAGC ACTCTACCAA GCTGAGCTAC TTCCCGAGTT 6240

AAATAGAAAA ATGCACCCTA GAGGAGTCGA ACCTCTAACC GCCTGATTCG TAGTCAGGTA 6300

CTCTATCCAG TTGAGCTAAG GGTGCTCCAT ATTATGCCGA GGACCGGAAT CGAACCGGTA 6360

CGATCGTTAC CAATCGCAGG ATTTTAAGTC CTGTGCGTCT GCCAGTTCCG CCACCCCGGC 6420

CTCTCTAAGC GAACGACGGG ATTCGAACCC GCGACCCCCA CCTTGGCAAG GTGGTGTTCT 6480

ACCACTGAAC TACGTTCGCA CTGTTTTCTT CTATCTAAAA ATGCCGGCTA CATGACTTGA 6540

ACACGCGACC CTCTGATTAC AAATCAGATG CTCTACCAAC TGAGCTAAGC CGGCTCATTT 6600

GTTATATCTT AATGCGGGTT AAGGGACTTG AACCCCCACG CCGTTAAGCG CCAGATCCTA 6660

AATCTGGTGC GTCTGCCAAT TCCGCCAAAC CCGCATATAT GACCCGTACT GGGCTCGAAC 6720

CAGTGACCCA TTGATTAAAA GTCAATTGCT CTACCAACTG AGCTAACGAG TCTAAAATAA 6780

cTTGCGTTAC CTTAAACGGT CCCGACGGGA ATCGAACCCG CGATCTcGCC GTGACAAGGC 6840

GACGTG 6846

(2) INFORMATION FOR SEQ ID NO: 199:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2911 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

GAATTCATTT TAAATAAAGA TACGGGAGAG GTAAGTGAAT TAAAACCTCA TAGGGTAACT 60

GTGACCATTC AAAATGGAAA AGAAATGAGT TCAACGATAG TGTCGGAAGA AGATTTTATT 120

TTACCTGTTT ATAAGGGTGA ATTAGAAAAA GGATACCAAT TTGATGGTTG GGAAATTTCT 180

GGTTTCGAAG GTAAAAAAGA CGCTGGCTAT GTTATTAATC TATCAAAAGA TACCTTTATA 240

AAACCTGTAT TCAAGAAAAT AGAGGAGAAA AAGGAGGAAG AAAATAAACC TACTTTTGAT 300

GTATCGAAAA AGAAAGATAA CCCACAAGTA AACCATAGTC AATTAAATGA AAGTCACAGA 360

AAAGAGGATT TACAAAGAGA AGAGCATTCA CAAAAATCTG ATTCAACTAA GGATGTTACA 420

GCTACAGTTC TTGATAAAAA CAATATCAGT AGTAAATCAA CTACTAACAA TCCTAATAAG 480

TTGCCAAAAA CTGGAACAGC AAGCGGAGCC CAGACACTAT TAGCTGCCGG AATAATGTTT 540

ATAGTAGGAA TTTTTCTTGG ATTGAAGAAA AAAAATCAAG ATTAAGATAA AAGCTATAGA 600

AAAAAATGGT TTATGTACTG AGATTAGATA GTGAGGTGAT GACATAGTTT TGTGAAAATA 660

GCCATTTATA ACTCAATTAT TTAGTTTACT TTACTTTACT AGTGATACTA TTTGGAGTTA 720

TTAATGGACT TAGTTTATAT AACTAATGAA TTGATTGAAA GGGTTAGTAT TGACAATATT 780

GGTCATATTG ACTAGAAAAT AGAGTCTATC AAAATTTAAA GGCTAATAGA GGTGATGAGA 840

CAATTTCGGC TCTTTGTCAA CTGTAGTGGG TTGAAGTCAG CTAAGCTCGA GAAAGGACAA 900

ATTTTGTCCT TTCTTTTTTG ATATTCAGAG CGATAAAAAT CCGTTTTTTG AAGTTTTCAA 960

AGTTTCGAAA ACCAAAGGCA TTGCGCTTGA TAAGTTTGAT GAGATTATTG GTCGCTTCCA 1020

GTTTGGCATT AGAATAGTGT AGTTGAAGGG CATTGACAAT CTTCTCTTTA TCTTTGAGGA 1080

AGGTTTTAGA GGATGAACTT GATTCAGATT GTCCTCAATG AGTCCGAAAA ATTTGTCAGG 1140

CTCCTTATTC TGAAAGTGAA AAAGCAAGAG TTGATAGAGA TTATAGTGGT GTTTCAAGTC 1200

TTCTGAATAG CTCAAAAGTT TATCTATAGT AGATTGAAAC TAGAATAGTA CACCTCTGCT 1260

TCTAAAACAT TGTTAGAAAT CGATTTGACT GTCCTGAATG ATTTGTCCTG TTATTATTTC 1320

ATTTTACTAT AAATCCACGT TTACGAATCT CTTTCCACAC TTGTTCAATG GGGTTCATCT 1380

CTGGTGTGTA TGGAGGAATA AATGCAAAAC CAATATTAGT CGGAATCTTT AAGGTACTTG 1440

ATTTATGCCA TATAGCATTG TCCATAACGA GTAAAAGATA ATCATCTGGA TAAGCTTGTG 1500

AAAGCTCCTA TTCCTAAAGC CCCTTTATAA CCTCTTGCGA GAGAGACTAT TGACTCAGCC 1560

CTTACTTCAT GCGGATGAAA CTTCTTATCG GGTTCTAGAG AGTCATAGCC ATCTGACCTA 1620

CTATTGGACC TTTTTGTCTG GGAAAGTTGA GAATCAAGCA ATCACGCTGT ACCATCATGA 1680

TCAGAGTCGG AGTGGTTCGG TAGTACAAGA ATTCCTAGGA GATTATTCTG GCTATGTTCA 1740

TTGTGATATG TTGCGGCAGT AACTTAGGAC TTTAGTCCTC TAGTTCTGCC TATGCGATAG 1800

CAGTCCAAGG TTTAGGAGCA AGGCGACGCT AAGCTTGGTA AACTGCGAAC CGCTAGAAGC 1860

TTATCGTCAA CTGGAAGAAG CTGAACTTGT TGGATGTTGG GCGCATGTGA GAAGGAAATT 1920

TTTTGAAGCG ACCCCCAAGC AAGCAGATAA ATCATCCTTA GGAGCTAAAG GTTTAGCTTA 1980

TTGTGATCAG TTATTTTCCT TGGAAAkAGA CTGGGAGGCT TTGCCAGCTG ATGAACGACT 2040

ACAGAAACGT CAAGAACATC TCCAGCCCCT AATGGAAGAC TTCTTTGCTT GGTGCCGCCG 2100

TCAGTCAGTT TTAGCAGGTT CAAAACTAGG AAGGGCAATT GAATACAGCC TCAAGTATGA 2160

AGAAACCTTT AAGACTATTT TGAAAGACGG ACATCTGGTC CTTTCCAATA ATCTAGCTGA 2220

ACGCGCCATT AAATCATTGG TTATGGGACG GAGTAAAAGA GTCCAGTGGA CTCTTTTAGC 2280

CTGAGCTCAG TTTAAAAAAG CGAGGGTGGT TATTTTCTCA AAGTTTTGAA GGAGCTAAAG 2340

CAAGAGCTAT TGTTATGAGC TTGTTGGAAA CAGCTAAACG TCATCAATTA TAGTGCGTTG 2400

AATCTATAAC AGTACGCATC GACTGCTAAA ACATTTCTAT AAATCAATTT TCCTTTCCTA 2460

ATCGATTTGT TCATATCTTA TTTCAATCCA TTATAAATAG CGAGAAATAT CTATCCTATC 2520

TTCTAGAATG TCTTCCAAAC GAGGAAACTC TCGTAAACAA AGAGGTTTTA GAGGTTTATT 2580

TACCATGGAC TAAAGTTGTA CAAGAAAAGT GCAAATAAGA AATCTCCAGA TTAGGAACTA 2640

TCCGTGAGTT CACTAATCTG GAGATTTTTC AATAGAtTCG TTATTGGGCG GTTACGATAT 2700

GATcACTACT TCGTCAGTCT TATCTACAAC CTCAAAACAG TGTTTTGAGC AACCTGCGAC 2760

TAGCTTCCTA GTTTACTCTT TGATTTTCAT TGAATATTAG AACAGAAAAA ATGCTTGGAG 2820

TATTTGTTTG TGTGTTTATT TTTATATAAC AAACTATAAA CAAAATAAAA ATATAAAAAA 2880

AGAGACAAAA AAGAACAGAA AGTAATTGAC A 2911

(2) INFORMATION FOR SEQ ID NO: 200:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6854 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

GAAAATAAGT CTTGACAGAA AGCGCTATCA ATGATAGAAT GAATTCAGAT AAAAAGATTT 60

ATTTTTAAAA CAAAAATGAA ACGTTTCAAA AAAAGAAATA AAGAGACAGC GCCAAGCGCT 120

ATCTTTTCTA GAAAAAAATG AAACGTTTCA AAAAAGGAGG TTGCTATGAA TAGCAAAGCG 180

AAGCAAGTTT CTCTTTGGGA AAGAATCAAG AAACAAAAAC TCTTGTTATT GATGACTGTC 240

CCCGGTTTAG TTTTAACCTT TATCTTTAAA TACATCCCTA TGTATGGGGT TTTAATCGCA 300

TTTAAAGATT ACAATCCTTT AAAAGGAATT TTAGGGAGTG ATTGGATTGG TTTTTCTGAG 360

TTTACAAAAT TCATATCCTC TCCCAACTTT GGTATCTTGT TAGCCAACAC ATTAAAATTA 420

AGTATCTATG GTTTATTGCT TGGCTTTTTA CCACCAATCA TTCTCGCGAT TATGCTCAAT 480

CAACTCTTGA GTGAAAAAGT CAAAAAACGA ATTCAGCTCA TTTTATACGC ACCAAACTTT 540

ATCTCAGTCG TTGTTATTGT CGGTATGATT TTCCTCTTCT TTTCAGTGGG AGGACCAATC 600

AACAATTTTC TTTCTATGTT TGGAATGAAG GCTGACTTCT TGACAAATCC AGACTTCTTT 660

AGACCTTTAT ACATCTTTAG TGGTATCTGG CAAGGAATGG GCTGGGCTTC AACGCTCTAC 720

ACGGCAACAT TGGTAAATGT AGATCCAGCC TTAGTAGAAG CAGCCCGACT GGATGGAGCC 780

AATATCTTCC AACGAATCTG GCACATTGAT ATTCCAGCTC TTAAGCCTAT TATGGTTATC 840

CAATTTGTTT TAGCTGCAGG TGGAATTATG AATGTCGGAT ATGAAAAAGC ATTCTTGATG 900

CAGACATCGT TAAATTTGCC AACTTCTGAA ATTATCTCGA CATATGTCTA TAAAGTTGGT 960

CTTGTATCAG GAGACTATTC TTACTCAACA GCGGTTGGTT TGTTTAATGC AGTGATTAAC 1020

GTAGTATTGC TTGTTGCAGT TAACCAAATC GTTAAACGCA TGAATAATGG TGAAGGAATT 1080

TAAGGAGGAA AGTATGAAAA ATTCGATTAT GGATACAAAA TTTGATAGAC GTATCTTACT 1140

CTTAAATAAA ATCATTATTG TCTTTATCGT TTTGATGACT TTGCTTCCTT TACTTTATAT 1200

CGTCGTAGCA TCCTTTATGG ATCCTAAGGT TCTGGTTAGT AGAGGGATTA GCTTTAATCC 1260

AGCCGATTGG ACTGTAGAAG GTTACCAGCG TGTATTCAGT GACCAATCTA TTCTAAGAGG 1320

TTTTATCAAT TCTCTACTAT ACTCTTTTGG ATTTGCAGCT TTAACAGTCT TGCTATCTGT 1380

GTTTACAGCT TATCCTCTTT CTAAGAAAGA CTTGGTTGGA CGTCGTTGGA TTAACTACTT 1440

CTTGATTGTA ACTATGTTCT TTGGTGGTGG TTTAGTCCCA ACTTACTTGC TCGTAAAAGA 1500

ATTGGGAATG CTCAATACTC CATGGGCTAT CATTGTTCCA GGTGCTGTTA ACGTTTGGAA 1560

TATTATTCTT GCTAGGGCCT ATTTCCAAGG ATTGCCTGAA GAATTAGTTG AAGCTGCTGT 1620

CATTGATGGT GCAAATGATT TACAGATTTT CTTCAAAATC ATGCTTCCTC TTGCAAAACC 1680

AATTATGTTT GTTCTCTTCC TTTATGCTTT TGTAGGACAG TGGAACTCAT ACTTTGATGC 1740

AATGATTTAT ATCAAGGATC CAAACTTGGA ACCATTGCAA CTTGTACTTC GTAAAATTCT 1800

CATTCAGAGC CAACCAGGTC AAGACATGAT TGGAGCACAA GCGGCTATGA ATGAAATGAA 1860

ACGTTTAGCT GAATTGATTA AATACGCAAC TATTGTCATT TCCAGCTTGC CATTGATTGT 1920

TATGTATCCA TTCTTCCAAA AATACTTTGA TAAAGGAATT ATGGCTGGTT CACTTAAAGG 1980

ATAAAAAAAG AAAAAATAAA AGGAGTTTTC TCATGAAATT CAAAACATTC TCAAAATCAG 2040

CAGTTTTGTT GACAGCTAGT TTAGCAGTAC TTGCAGCCTG TGGCTCAAAA AATACAGCTT 2100

CAAGTCCAGA TTATAAGTTG GAAGGTGTAA CATTCCCGCT TCAAGAAAAG AAAACATTGA 2160

AGTTTATGAC AGCCAGTTCA CCGTTATCTC CTAAAGACCC AAATGAAAAG TTAATTTTGC 2220

AACGTTTGGA GAAGGAAACT GGCGTTCATA TTGACTGGAC CAACTACCAA TCCGACTTTG 2280

CAGAAAAACG TAACTTGGAT ATTTCTAGTG GTGATTTACC AGATGCTATC CACAACGACG 2340

GAGCTTCAGA TGTGGACTTG ATGAACTGGG CTAAAAAAGG TGTTATTATT CCAGTTGAAG 2400

ATTTGATTGA TAAATACATG CCAAATCTTA AGAAAATTTT GGATGAGAAA CCAGAGTACA 2460

AGGCCTTGAT GACAGCACCT GATGGGCACA TTTACTCATT TCCATGGATT GAAGAGCTTG 2520

GAGATGGTAA AGAGTCTATT CACAGTGTCA ACGATATGGC TTGGATTAAC AAAGATTGGC 2580

TTAAGAAACT TGGTCTTGAA ATGCCAAAAA CTACTGATGA TTTGATTAAA GTCCTAGAAG 2640

CTTTCAAAAA CGGGGATCCA AATGGAAATG GAGAGGCTGA TGAAATTCCA TTTTCATTTA 2700

TTAGTGGTAA CGGAAACGAA GATTTTAAAT TCCTATTTGC TGCATTTGGT ATAGGGGATA 2760

ACGATGATCA TTTAGTAGTA GGAAATGATG GCAAAGTTGA CTTCACAGCA GATAACGATA 2820

ACTATAAAGA AGGTGTCAAA TTTATCCGTC AATTGCAAGA AAAAGGCCTG ATTGATAAAG 2880

AAGCTTTCGA ACATGATTGG AATAGTTACA TTGCTAAAGG TCATGATCAG AAATTTGGTG 2940

TTTACTTTAC ATGGGATAAG AATAATGTTA CTGGAAGTAA CGAAAGTTAT GATGTTTTAC 3000

CAGTACTTGC TGGACCAAGT GGTCAAAAAC ACGTAGCTCG TACAAACGGT ATGGGATTTG 3060

CACGTGACAA GATGGTTATT ACCAGTGTAA ACAAAAACCT AGAATTGACA GCTAAATGGA 3120

TTGATGCACA ATACGCTCCA CTCCAATCTG TGCAAAATAA CTGGGGAACT TACGGAGATG 3180

ACAAACAACA AAACATCTTT GAATTGGATC AAGCGTCAAA TAGTCTAAAA CACTTACCAC 3240

TAAACGGAAC TGCACCAGCA GAACTTCGTC AAAAGACTGA AGTAGGAGGA CCACTAGCTA 3300

TCCTAGATTC ATACTATGGT AAAGTAACAA CCATGCCTGA TGATGCCAAA TGGCGTTTGG 3360

ATCTTATCAA AGAATATTAT GTTCCTTACA TGAGCAATGT CAATAACTAT CCAAGAGTCT 3420

TTATGACACA GGAAGATTTG GACAAGATTG CCCATATCGA AGCAGATATG AATGACTATA 3480

TCTACCGTAA ACGTGCTGAA TGGATTGTAA ATGGCAATAT TGATACTGAG TGGGATGATT 3540

ACAAGAAAGA ACTTGAAAAA TACGGACTTT CTGATTACCT CGCTATTAAA CAAAAATACT 3600

ACGACCAATA CCAAGCAAAC AAAAACTAGA GGTTGATTAT GGGAGATAAG AAATACACAG 3660

TAGAAAAAGC CAATCGTTTT ATAGCAGAAA ATAAACATCT CGTTAATACT CAATATAAGC 3720

CTGAAGAACA TTTTTCAGCT GAGATTGGTT GGATCAATGA TCCAAATGGA TTTGTCTATT 3780

TTCGTGGAGA ATACCATCTC TTTTATCAAT TCTATCCATA TGATAGTGTT TGGGGGCCTA 3840

TGCACTGGGG ACATGCTAAA AGTAAGGACT TGGTGACTTG GGAGCACTTG CCAGTGGCAC 3900

TTGCTCCTGA CCAAGATTAT GACCGAAATG GTTGTTTCTC AGGCTCTGCC ATTGTCAAGG 3960

ATGATCGCCT CTGGCTCATG TACACTGGAC ATATCGAAGA AGAAACCGGT GTCCGCCAAG 4020

TGCAAAATAT GGTATTTTCA GATGACGGGA TTCACTTTGA AAAGATTTCC CAAAATCCAG 4080

TTGCAACTGG ATCAGACTTA CCAGATGAGT TGATTGCTGC TGATTTCCGT GATCCAAAAC 4140

TCTTTGAAAA AGATGGACGC TATTACTCCG TAGTAGCTGC CAAACACAAG GATAATGTGG 4200

GCTGTATCGT TCTACTAGGG TCCGATAACC TAGTAGAATG GCAGTTCGAA TCCATCTTTT 4260

TAAAAGGGGG AGAACACCAA GGTTTTATGT GGGAATGCCC AGATTACTTC GAGTTAGATG 4320

GGAAAGATTG CCTTATTATG TCACCCATGC GTTATCAGCG TGAGGGAGAC TCATATCATA 4380

ACATCAACTC ATCGCTTTTG TTCACGGGTA AGGTAGATTG GAGAGAAAAA CGTTTTATCC 4440

CAGAATCAGT TCAAGAAATT GATCATGGCC AAGACTTCTA TGCGCCTCAA ACATTGTTGG 4500

ACGATCAAAA TCGTCGTATC CTGATTGCTT GGATGCAGAC ATGGGGGCGT ACCCTTCCAA 4560

CCCATGACCA AGAACACAAG TGGGCATGTG CCATGACTCT ACCTAGAATT CTAAGATTGG 4620

AAGATGGCAA ACTAAGACAA TTCCCTGTTA AAAAAGGCCA ATATCAAATC CAAATAGATA 4680

AAGATTGTCA TTACCACTTA GGAAATGATA TAGATTATCT TGAATTTGGT TATGACAGTA 4740

ATGCGCAGCA AGTTTACATT GATCGTAGCC ATCTTATTCA AAAAATTCTA GGTGAAGAAG 4800

AACAGGACAC TAGTCGACGG TATGTAGATA TTGAAGCTAA AGAATTGGAA GTTGTTCTAG 4860

ATAAAAATTC CATCGAGATT TTTGTCAATC AAGGTGAAGC AAGCTTGACT GCAACTTATT 4920

ACTTAACGGT GCCAGCTGAG CTATCACGAA TTGATTAAAA ATTAAGTTAT TTCTCCTAAA 4980

GAAAAAGTTC TCTTTCTAAA ATAGTGGAAA GAGGACTTTT TGTGTTTTGG GTATATAAGC 5040

TTAGTTTATG GTATTTGTAA AATTGGTGTT GGATTATGAT TTAAGCTAGT TTTCTAAAGA 5100

ATTTGAAAAA AATTTTATTT AAGCAAAAAA ACCTTGGTTC CAAGGCTTTT CCTGTTGTAT 5160

TTAGATGCCC CCTACAGGGA TTGTAGGAGA TATGTTGCTT AGATGTTCTT GATTTTCTGG 5220

TGTTTTGTAA CGTTTAAATG AGTTTTTTGA GTTTGTTGGT GGGGCGTTGC CCGGCAATTG 5280

CCCGACTTAT TGCTTGAAAA AGAATTTAAA ATATAGTATA GTTAATTATA GATTAACACT 5340

TGCTTGGAGG AACTGATGAA GAACAATGAA AGATTAGGTA TTAAATTAAG TAGAGATAGC 5400

GTTTTAGGAT TGAGGGAAGT TAGAAGGCTT TATTTAGGCA GTTCAGATAT CCCAGTTTCT 5460

GATGGCTATG TGATTGAAGT TGCTTATAAC CAGATATCAC ATGAGATTGA TATTATTGAT 5520

TGGGTAGAGT TGAACAAGTC AAAAATTAAG ATAAGTGAAA TTAGTGAAAG CGTGGATATA 5580

GATGCCACTA GCTTGAGAAC AACTTTGACT TTAGACACAT TAGTATATGA AGGTATGAGA 5640

GATATACAGT TAAAGTTGAG AGAGCTTACA AAGGGGAGAG TATTCTTTTC ATTTGTAGTG 5700

AAGTTAGTTT TGTTTGCTTC TATTTTAAAG AAAAAAGATT TACTAGAAAA ATTTCAAGAA 5760

AAGTGTTAAT CAAGTATTGA CACTTTATCT GGATTTCGGT ATAATATGCT TAGAAAGGAA 5820

TCTTTCTAAA TTTTTTTCGT CCTTATGTGT TAATCAAAGA CGAATACAAA AACATATTTT 5880

TTTACTCTAA AAAGTGTTAA TCAATGATGT ATTTGTTAGA GAGGTAGATA AATGGAATTG 5940

AGAGCACCAC CAGTTATAAT AGTATAAAAC GTATAATAAA AATATTTTAA CTTGAATTAT 6000

AGAAAAGGAG AAACAAATCA TGAAACAAAA ACAACCGATT GTTTCTAGAA CGAAACAACA 6060

TACATTTGAA GAGCTTATTC AAGACCAAAA GTTAGAAAGA TTGGCTAAGT TGTCGCCCGA 6120

TTTGGTTGGA AGGTATGGTT TTACTGCTAG CTGTGCGTCT TCATTTGCGA ACTTGATTAA 6180

AGAAGCGTAT GGGGGTAAAA ATCTAAACGT AGTTTATGCG AGTCGGATGT TGGCTCTCTG 6240

GAATATTGCT TGCAGTTGTT ATCATAAGGC TGATGGGTAT TCTTTAGCAG ATGCGCTTTT 6300

TAGTGATAAA AAAATTTGTC TAGATTCTTA CTATTACCAC AAGAATACCT CTAATACCAT 6360

AACTAGTGAT GTGATAAAAG ATGTTTACGA TAATTATAAT AATTATATGG TTTTAACTCG 6420

AGAAGCGACA CCTGAATACA TTTATGTTGT ACAAACTGAA ATGCCAAAAG ATTCAGATTT 6480

ATATTTTTAT ATTAGAGAAG TTCTGGGATT ATCGTTTAGT ACCATGCATT ATGCATTTTT 6540

AGTCAAGGTT CTTGCAGGAG CGCTTGCTAG AAAATATAAG CCATATCGAA ATTGAATTAT 6600

TTAAATTTAT ACTCTTCGAA AATCAAATTC AAACCAAGTC AGCTTCGCCT TGCTGTACTC 6660

AAGTGCTGTC TGTGGCTAGC TTCTTAGTTT GCTTTTTGAT TTTCATTGAG TATTACTCTT 6720

ATGGTAGTTA TTTATGGCAT AATAATATTG ATTTGGGAGT TATAGCGAAA ATTTTAGGTT 6780

CTATAATATT TGTAGTGGGT AAACCACTAT AGATATTATG GAGCCTATTT ATTGTAGAAA 6840

AAAGTCCCAT ATGA 6854

(2) INFORMATION FOR SEQ ID NO: 201:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3895 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

TCCTTGCTAA GTTTATACTC AATGAAAATC AAAGAACAAA CTAGGAAGCT AGCCACAGGT 60

TGCTCAAAGC ACCGCTTTGA GGTTGCAGAT AAAACTGACA CGGTTTGAAG AGATTTTCGA 120

AGAGTATTAA TTTACATAAA TAGCCAGTGT TTGATAGGGT TTGAGTAGAA TTTTCTCAGA 180

1180

CACTTCTGCA TCTTCATAGT TTGATATCAA AATCTGTCCA TTTTGGTAGA CTGCTGGCAA 240

GTCGATTTcA CTTCTTTAGC ATAAAAGTTA TTGAGCACTA GTAACTTTTG ATCCTCAAAC 300

TGGCGTTCAA AAGCGTAGAC TTGTTTGCTA TCTTCAAAGG CTGGTTTGTA ACTTCCTTCT 360

GAAATGATTG GCATTTCCTT ACGCATCGAA TCAAGTCTTG ATAGAAGGTA AAAATCGGAC 420

CCTGGATTTC ATTTTCTACA TTGATGTATT TATAGGATTT ACCAGCTTTC AACCAAGGAG 480

TGCCTGTTGA AAATCCTGCA TTTTCCGAAG CATCCCACTG CATGGGAATG CGTGAATTAT 540

CACGCGACTT AGCTTGAATA ATCTGGAAGG CTTCTTGCTG ACTCTTTCCT TCTTCTAAGA 600

GCATCTGATA GGCATTAAGC GATTCGACAT CCACATAATC AGCCATAGAA TCATAGTCTG 660

GGTCAATCAT CCCGATTTCC TCACCCATGT AGATATAAGG TGTCCCACGT GACAGGTGAA 720

TGCTGGCTGC TAGCATGGTG GCTCCTTCCT TGCGGAAGTT TTGAATATCG ACAAAACGGT 780

TCAAGGCACG TGGTTGATCG TGATTATTCC AAAAGAGGGC ACTCCAACCG TCTTTATCAC 840

TCATTTCCTT ACCCCAACTA TGGTAAAGAC TCTTCAACTC TTCAAAATCA AAGGGAGCCA 900

AGGTCCACTT TTGTCCATCC TTATAGTCCA CCTTGAGGTG ATGAAAATTA AAGGTCATGG 960

ATAATTCCTG ACGATCAGGC GACGAATAGA GGACACAGTT TTCCATGGTG GTAGAAGACA 1020

TTTCCCCAAC TGTCATAAAG CTATCGTCGG ATCCAAAAGT GGCTTGGTTC ATCATACGCA 1080

AATAGTTATG AACGATGGGT TTGTCTGTAT AAGCTGGCTT CCCTTCATTT TCAGGACAGT 1140

CCACTGAAAC CTCGTCCTTA CCGATCAAAT TGATCACATC AAATCGGAAA CCTTTGACAC 1200

CCTTGTCGCG CCAGAAATTA ACAACCTTGA AAAGCTCCTT ACGGACATTG GAATTGCGCC 1260

AGTTAAGGTC AGCCTGGGTC TCATCAAATA GGTGAAGATA GTATTTCCCA GTATCCCCGA 1320

AAGGCGTCCA TGCAGAACCA CCAAACTTAG ACTGCCAATC TGTTGGTTGG TCTTGGATGA 1380

AGAAAAAGTC TTGATAATAC TTATCACCAG CTAGGGCTTT CTGAAACCAT TCATGCTCTG 1440

TCGAACAATG ATTAAGTACC ATGTCCAGCA TAAAGTCAAT CTTGTGCTCT TTACCGACAC 1500

ACACCATTTT CTCAAAATCA GCCATATCAC CAAAAAGAGG ATCCACTGCC ATATAATCTG 1560

AAATATCGTA ACCATTATCC CGTTGAGGGC TTGGATAGAA TGGATTGAGC CAGACCATAT 1620

CCACACCTAG TTTGGCTAAA TAGGGAATTT TTTCGATAAT CCCACGGAAA TCCCCAATAC 1680

CGTTTTCAGT GGTGTCTTTG TAAGATTTTG GATAGATTTG ATAGACTACT TTTCCTTTAT 1740

CAAGTGTCAT CTGTTTCTCC TTTTCTGATA AAAGGGAGGA AGCAGTCTTC CGTCCCTATT 1800

TGTGCTATTT CAATTATACT CAATGAAAAT CAAAGAACAA ACTAGGAAGC TAGCCACAGG 1860

TTGCTCAAAA CACTATTTTG AGGTTGCAGA TAGAGCTGAC GTGGTTTGAA GAGATTTTCG 1920

AAGAGTATTA GATTCGTGTA GCGACCATGA GAGATGCTCC AGCTTGGATC GTTGTCGGAT 1980

AAGTTCCGGG AATAGTCGCT GTATAAGCAT CTTGGTTGGT GATGATAACA GGAGTTTCTG 2040

TCACCAGACC TGCAGCCTTA ATGACATCCA TATCAAAACG AATCAGTTGC TGACCAACTG 2100

TAACGTGATC TCCTTGGACT ACAAGACTTT CAAAACCTTT GCCATCAAGA CCTACTGTAT 2160

CCATACCGAT GTGGATGAGC AATTCAACTC CCTCGTCAGA GACAATGCCG ATGGCATGCT 2220

TGGTAGGGAA AAGAACCGTC ACTGTCCCAT TAACTGGAGA GGTCAACTCA CCTTGGCTTG 2280

GTTCAATGAC TAGACCTTGC CCCATGACAC CTGATGCAAA AATAGGATCC GTCGCTTGAC 2340

TCAATTCTTT CACTTGGCCA GTTAGTGGGC TGATAATTTC TACCGAAGTA AGTTCTACTG 2400

GTTCATGGTT CACAAATTCT GCTTCTTCTT GAGCAACGAA TTCTGCCTGC AAGTTCGTAT 2460

CGCCCTCTGT TTTTGTAAAG AGACCAGCCT TGCGGAAGAA GAAAGTCAAG AGCATTGGAA 2520

CAACAATCGC AACTAGCATA GTTCCTGCAA ATGGCAGCAT GTATTGAGGT TGAATAGAGA 2580

GAATACCTGG CAAACCACCG ATACCAATAG AAGCCGCAGT TACATTAAAA GTAACGGATA 2640

ACATGCCTGC AAGGGCTGAA CCAGTCATCC CAGCAACAAA TGGATAAATA TATTTTACGT 2700

TAACCCCAAA AAGAGCTGGT TCTGTAACAC CGAGATAGGC TGAAATGGTT GCAGGAAGTG 2760

AAACCTGAGC CTCACGCTCA TCATGGCGAT GCATGAAATA ATAGGCAAAC ACGGCTGAGC 2820

CTTGAGCAAT ATTAGAAAGA GCAATCATTG GCCATAGGGC AGTGCCACCA GCATCCGCAA 2880

TCAATTGTGT ATCAATGGCA TTGGTCATAT GGTGCAGACC TGTGATGACA AATGGAGCGT 2940

AGAGGGCGCC AAAAATTGCA CCGAAGAGCC ATTTAACTGG ACCAGTTAAA CCTGCCAAGA 3000

CAACTGATGA AAGTCCTTGT CCAATTGTCC AACCGATTGG TCCCAAAACA GTATGAGCCA 3060

AAATCAAGGC TGGAATCAAT GACAAGAAAG GTACAAAAAT CATAGAAATG ACTTCTGGGA 3120

TATGCTTGTG CCAGAAGATT TCAAGATAAG ACAGACTCAA ACCTGCAAGC AAGGCTGGGA 3180

TAACTTGGGC TTGGTAACCG ATACGATTAA CAGTAAAATA GCCAAAATTC CAAACCCAGT 3240

TTGCCGCGAT ATCAGCTGCT GGCGTTGAAG CAACCGCATA GGCATTGAGC AACTGAGGCG 3300

ATACCAAACA GATTCCGAGA ACAATTCCCA AAATTTGGCT GGTTCCCATC TTACGAGAAA 3360

CAGACCAAGT AATCCCTACT GGTAAGAACT GGAAGATAGC TTCACCAGGC AACCAGAGGA 3420

AGTGATTGAC ACCTGCCCAA AACTGAGAGG ATTCTGTGAT GGTCTTGCCA TCCAACATCG 3480

ACCAATGGAC ACCTTCCAAG ACATTACGGA AACCGAGGAT CAATCCTCCG ACTATCAAGG 3540

CTGGAATAAT CGGAGTAAAA ATCTCCGCCA GAGTGGTCAT AACACCTTGG ACCACGTTTT 3600

GATTACTCTT AGCTGCAGAC TTGGCTGCTT CTTTGGAAAC ACCCTCAATA CCTGAAACGG 3660

CTGTAAAATC ATTATAAAAG ATGGGCACGT CATTTCCAAT GATTACCTGA AATTGACCTG 3720

CATTTGTAAA GGTTCCTTTA ACAGCTGGAA TTGACTCGAT AGCTTTAACA TTAGCCTTCT 3780

TATCATCTCC TAAAACAAAC CGCATCCGTG TCGCACAGTG AGTTACGGCA GTCACATTTT 3840

CTTTGCCTCC GATTGCCTGA AGCAGATCTT TGGCTTCTTG TTCAAATTTT CCCGG 3895

(2) INFORMATION FOR SEQ ID NO: 202:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3936 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

AGGATCGCCG CTCCAGCTAC TAAGTCTCGT GCAGTGCCGA TTTATCAAAC AACATTTTTT 60

GTTTTTGATG ACACGTAGGA AGGTGCCGAT CTGTTTGCCT TGAGGAAACC AGGGAACATT 120

TATACTCGTA TCACCAATCC TACAACAGCT GCCCTTGAAG GTGGTGTTGA AGCGCTAgcA 180

ACAGCATCAG GTATGACTGC AGTGACTTAT ACGATTTTGG CGATTGCCCA TGCTGGTGAC 240

CATGTAGTGG CTGCTTCGAC TATTTACGGT GGAACCTTCA ATCTTTTGAA AGAACCCCTT 300

CCTCGTTATG GTATCACAAC AACCTTTTTC GATATTGATA ATTTGGAGGA AGTAGAAGCA 360

GCTATCAAAG ACAATACCAA GCTTGTCTTG ATTGAAACCT TGGGTAACCC CTTGATTAAT 420

ATTCCAGACC TGGAAAAACT GGCAGAGATT GCTCATAAAC ATCAAATCCC ACTTGTGTCA 480

GACAATACTT TTGCAACACC TTATTTGATT AACGTCTTCT CTCATGGCGT TGACATTGCC 540

ATTCACTCTG TGACTAAGTT TATCGGTGGG CATGGTACAA CTATTGGAGG AATAATTGTC 600

GATAGTGGTC GTTTTGACTG GACGGCTTCA GGGAAATTCC CTCAATTTGT TGACGAGGGT 660

CCAAGCTGCC ACAATTTGAG CTATACTCGT GATGTGGGTG CAGCAGCCTT TATTATAGCT 720

GTTCGAGTTC AATTGCTTCG TGATACAGGT GCAGCCTTGT CACCATTCAA TGCTTTCCTC 780

TTGCTACAAA GACTTGAAAC CTCTTCACTT CGTGTGGAAC GCCATGTACA AAATGCTGAG 840

ACAATTGTTG ATTTTCTTGT CAACCATCCT AAGGTAGAAA AGGTAAATTA TCCAAAACTT 900

GCAGATAGTC CTTATCATGC CTTGGCTGAG AAATACTTGC CAAAAGGTGT CGGTTCAATC 960

TTTACCTTCC ACGTCAAAGG TGGCGAGGAA GAAGCACGCA AGGTCATTGA TAATTTAGAA 1020

ATCTTTTCTG ACCTTGCAAA CGCGGCAGAT GCTAAATCGC TTGTTGTCCA TCCAGCAACA 1080

ACCACTCACG GTCAATTGTC AGAAAAAGAC CTAGAAGCAG CAGGTGTCAC ACCAAACTAA 1140

ATTCGTTTGT CAATCGGTCT TGAAAATGTA GAAGATTTGA TTGAAGACTT GCGCTTGGCC 1200

TTGGAAAAAA TTTAAAGTAA AAGAAGATAA ACAGTGGGCT TCGACTCACT GTTTTTGATT 1260

TTCCCTCAGG CATGATATAA TGGTTACAGA AGTCTAGAAA GAGGAACGAT ATGAACGAAA 1320

TCAAATGTCC CAACTGTGGG GAAGTCTTTA CAGTAAATGA GAGTCAGTAT GCCGAACTCT 1380

TGTCCCAAGT GAGAACGGCA GAGTTTGATA AGGAACTACA CGATAGGATG AAGCAGGAAC 1440

TGGCCTTGGC TGAGCAAAAG GCCATGAATG AGCAACAGAC TAAACTGGCT CAGAAGGATC 1500

AAGAAATTGC GCAATTACAG AGTCAGATCC AAAACTTTGA TACAGAAAAA GAATTGGCCA 1560

AGAAAGAGGT TGAACAGACA AGCCATGAGG CTCTCTTGGC TAAGGACAAG GAAGTACAGC 1620

TCTTAGAAAA TCAGTTGGCT ACCTTGCGTT TGGAGCATGA AAATCAACTA CAAAAGACCC 1680

TTTCTGACCT AGAAAAAGAA CGGGATCAGG TTAAAAACCA ACTACTTTTG CAGGAAAAGG 1740

AAAATGAATT ATCTTTGGCT TCTGTTAAGC AAAACTACGA AGCCCAGCTC AAGGCAGCTA 1800

GTGAACAAGT CGAGTTTTAT AAGAATTTTA AGGCTCAACA ATCTACAAAA GCGATTGGGG 1860

AAAGCCTAGA ACAGTATGCA GAGAGTGAGT TTAACAAGGT TCGTAGTTTC GCCTTTCCAA 1920

ATGCTTACTT TGAGAAGGAT AACAAGGTCT CTTCGCGTGG GTCTAAAGGG GACTTTATCT 1980

TCCGTGAGTG TGATGAAAAT GGAGTTGAAA TCATTTCTAT CATGTTTGAG ATGAAAAACG 2040

AAGCGGACGG AACAGAGAAG AAGCACAAGA ATGCAGATTT TTACAAGGAA TTGGACAAGG 2100

ACCGTCGGGA GAAGAACTGT GAGTATGCCG TTTTGGTGAC CATGCTTGAG GCTGATAATG 2160

ACTACTTTAA CACAGGGATT GTTGACGTCA GTCACGAGTA TGAAAAAATG TATGTTGTTC 2220

GTCCTCAATT CTTTATCCAA TTGATTGGTC TCTTACGTAA TGCGGCGCTA AATTCCCTAA 2280

AATACAAGCA GGAGTTGGCC TTGGTTCGCG AGCAAAATAT TGACATTACG CATTTTGAGG 2340

AAGATTTGGA TGCCTTTAAG CTAGCTTTTG CTAAGAACTA TAATTCAGCT TCGACTAACT 2400

TTGGAAAAGC TATTGATGAA ATCGACAAGG CCATCAAACG CATGGAAGAG GTTAAGAAAT 2460

TCCTGACCAC ATCTGAAAAC CAACTCCGTT TAGCTAACAA CAAATTGGAA GATGTCTCTG 2520

TTAAAAAATT GACCCGGAAA AATCCAACAA TGAAAGCGAA GTTCGAAGCA CTGAAGGGGG 2580

AGTAGAAAGC AAAAATGAAC GGTATTATTA ACTTAAAAAA GGAAGCAGGA ATGACCTCGC 2640

ATGATGCGGT TTTTAAACTG CGTAAGATTT TGGGAACCAA GAAAATTGGT CATGGTGGAA 2700

CCTTGGATCC GGATGTGGTG GGTGTTTTGC CGATTGCGGT TGGCAAGGCG ACACGCATGG 2760

TCGAGTTTAT GCAGGACGAG GGTAAGATCT ATGAGGGGGA AATCACTCTG GGCTATTCCA 2820

CGAAGACTGA GGATGCTAGT GGGGAAGTGG TCGCAGAAAC CCCTGTTTTG TCTCTCTTGG 2880

ATGAAAAGCT TGTTGATGAA GCGATTGCTA GCTTGACTGG GCCTATTACT CAGATTCCCC 2940

CTATGTATTC GGCAGTTAAG GTTAATGGTC GCAAGCTCTA TGAGTATGCG CGTGCTGGTC 3000

AGGAAGTGGA GCGTCCAGAA CGTCAGGTGA CCATTTATCA ATTTGAGCGA ACAAGTCCGA 3060

TTTCTTATGA TGGCCAACTT GCCCGATTCA CTTTTCGTGT AAAATGCAGT AAAGGGACGT 3120

ACATCCGTAC TTTGTCAGTT GATTTGGGTG AAAAGCTTGG TTATGCGGCT CATATGTCCC 3180

ATTTGACTCG TACTAGTGCT GCTGGCTTAC AATTAGAAGA CGCTCTTGCC TTGGAGGAAA 3240

TTGCTGAAAA AGTAGAGGCT GGGCAATTAG ATTTTCTCCA TCCTTTAGAG ATTGGGACAG 3300

GTGACCTTGT CAAAGTTTTC CTAAGTCCAG AAGAGGCTAC AGAAGTTCGC TTTGGTCGTT 3360

TTATTGAGCT AGACCAAACG GACAAAGAAC TGGCTGCCTT TGAAGATGAT AAATTGTTAG 3420

CCATTCTAGA AAAACGGGGC AATCTCTATA AGCCAAGGAA GGTTTTTAGC TAGATCGTTT 3480

AGGAATAAAA ATCGGGTGAT AGATAACAAT TGCTTGATAA AACCCCATAC TAATAGTAGA 3540

ATGGTTTTGG GAATTATAAT ATTCCAATTG TTGCGAGTTG TAGGTACTCA AATAATCTAT 3600

ATAGAAATTT AGAGGTGTGA AATGAAGCAA TTTAAAATTC TTTCAGATAA ATATTTAGAG 3660

TCCATTACAG GTTCTGATGG GAACTTAGGC CCAGGATTTG GTGTGATAAT TCCATGATGC 3720

GAAATGAGTT TCGAGAAAGG GTGGAGCAAC TTCTTCAACA AAAAGAAATA AATGAAAATA 3780

GTGAGTTGAG TCACCTGTTT CGTCTTGCTA TACAAAATTT AGACAGAAAT GAAAAATACC 3840

AATCGGTCAT GGCCAATTTG AGTCAAGGGT TGTCACTTTA CCTCATGACG CATCATTACC 3900

AGGCACCTAA GTCTGTCATT GATTTTGGTT TATGGA 3936

(2) INFORMATION FOR SEQ ID NO: 203:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3230 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

CATCCAGCAA CTGCTCCTCT GAGCGTTTCA AAATTGATGT AATTTTTCTA GTTTTTTCTA 60

ATAAATGTGC CATTTTTCAC CTCGAATTTA ATCGCTATCA TTATAACATA AAAACGTCTC 120

TTTTTCAATA ATTATCTGAA AATTCCTTAT TGACTTGCAT TGACTTACAA TTTAATTAAA 180

AACCAGAATA TTTTTAATTA AATTGTTCCT TTTCTATTGA CAAGTTGCCT ATTTTTGTGT 240

ATCATAATAT TATAAAAGAT AATATAATAA TTTTATTTGT CTTTTCACAT TCGGTCTCCT 300

TATATAAAAA AGCGATTCAT TTTGAACCGC TTTTTCTTAT TTATCGCCTT TGTTACGAAT 360

AACAAAGCCT GTTTGCTTTT CGCTTAAAGT ATTGCGTGGT TTTTTATTAT CCTTACGGTA 420

ACGTTTTTCC TTATCAAAAC GATCGTTGCC ACGACTTCCT TTTTTGAACT CATCACGGCG 480

ACCATTGCCA CGGCGATCAC GCTCTCGACG GTCGTCCCCA CGACGGCCTC CACGACCTCC 540

CTTAGCTTTA CCACCGAAAC CATTACCTGA TGGTTTAAAC GGTAGTGGtT TTTCACGTGC 600

AATCTCCACT TCTGGAAGGC TATCTGGGTC TTGGACTGTC AGACTCAAGA TATACATTGC 660

CAATTCTTCT GGAGTAAACT CAGCAGCCAA TTTGCGAGCA TCCTTACCAA ATTTCTCAAA 720

GTTGGCACGA ATGGTTTCAT CTGCAAAATC ACGTTCGATT TTCTTGAGAG CTACCTGTTT 780

TTTTGATTGG AAGGATTCTT CTACACTTGC AGGTTTGAGA CCTTTCATGC GTTTCTTAGT 840

CAAGTTTTCA ATGATTTGAA GGTAACCCAT TTCGTTTGGA GCAACAAAAG TAATAGATTG 900

ACCTGACTTA CCAGCACGAC CTGTACGACC GATACGGTGA ACATAACTCT CAGGATCTTG 960

TGGAATATCG TAGTTGTAGA CATGGGTCAC ACCTGAAATA TCCAAACCAC GCGCTGCAAC 1020

GTCTGTCGCA ACCAAAACAT CAAGATTGCC ATTTTTAAAG TCACGAAGGA CACGAAGACG 1080

TTTGTTTTGG TCTAGGTCGC CATGAATTCC TTCTGCACGG AAGCCACGAA TTTTCAAACC 1140

ACGAGTCAAT TCATCCACAC GGCGTTTGGT ACGACCAAAT ACAATAGCGA GTTCTGGTTG 1200

TGCCACATCC ATGAGACGAG TCATGGTGTC AAATTTTTCT TGTTCCTTAA CACGGATATA 1260

GTACTGGTCA ACCAATTCTG TTGTCAATTC CTTAGCCGCA ATCTTGACAT GTTCAGGGGC 1320

TTTCATAAAC TGAACACCGA TACGTTTGAT GGCATCTGGC ATAGTTGCTG AGAAAAGCAA 1380

AGTTTGACGG TTCTCAGGTA CACGGGAAAT AATGGCTTCG ATGTCTTCAA GGAAGCCCAT 1440

GTTAAGCATT TCATCCGCTT CGTCAAGGAT AAGGGTTTCA ATGTCTTGTA ATTTCAAGGC 1500

CTTGCGTTTA ATCAAGTCCA AGAGGCGACC TGGAGTTCCC ACCACAATAT GGGCACCAGA 1560

TTTAAGAGCC TTAATTTGTT TTTCAATGCT TGATCCGCCA TATACTGAAC GGACTTTGAC 1620

TCCCTTACTA CGACCAAAGC GGAAGAGTTC TTCTTGACTT TGGACAGCTA GTTCACGAGT 1680

TGGAGCGATG ACCAAGGCTT GGATAGTCGC TTCTTCTGTA CGGATTTTTT CAAGGGTAGG 1740

CAAGCCAAAG GCTGCAGTTT TTCCTGTACC AGTCTGAGCT TGACCGATAA CATCCTTGCC 1800

TTCAAGGGCC AAAGGAATAG TTTGTTCTTG GATAGGACTA GCTTCTACAA AACCAGCTTT 1860

TTCAATTTCT GCTAGCAAAT CAGCAGACAA GTTTAATTCA TTAAATTTCA CGTTATTCTT 1920

CTTTCTAAAG GTGGTGCGAA GCCACCCTAT AGGGCTTAGT TTATACTTTT CTTTTTATGA 1980

CGTATTTTCA TATAACTAGA TATAAAATCG TGTTGCTTCT TTTCCACAAA AGAAAAGTAC 2040

TGTTTTCTTT GCAACCTATC TAGTATAACA CAAGACCAGA GCAAAAGATA GCCCCATTTC 2100

TACAGAAAAT CATGTAAGCG CTTTTTGACT TTCTTTTTTG ATTGAACGAC CTAGATAATA 2160

AGACAAAGCC AAGGCGATAC TGTATAAAAT GAGAAAAACG AACAAGGTTT GTGTGTACGA 2220

ATGAGCCATT TTATAAGTCT CTGCTAATAA AATAGGTCCC GCTAAACCAG CCATTGCCCA 2280

AGCTGTTAAA ATATAACCAT GCAGAGCGGC CAATTCCTTG GTTCCAAAAA TATCACTGAG 2340

ATAAGCTGGA ATCAAAGAAA AACCAGCTCC ATAGCAAGTC ATCAAAATAG ACATAGCAAC 2400

TACAAATAAA ACGGAATCTG TAAAGAGCCA AAGTGAGAGA GAAAAGAAAA GATTGACAAG 2460

CAGTAATATA CTAAAGGTTA GAGGGCGACC GATATAGTCA GACAAACTCG CCCAGAGCAA 2520

GCGACCAAAT CCATTGAAAA TCCCCAAAAC ACCCACCATT ACTGCTGCAT GACTTGTAGA 2580

CAAGCCAGCC ATCTCCTGTG CCATTGGCGA TGCCGCTGAA ATTAAGCCTA AACCACAAGC 2640

TATGTTGATA AAGAAAATAA TCCAAAGCAT ATAAAACCGA TTGCTTTTTA GAGCCTGATT 2700

TGCAGCCATT CCTTGCGTCA AAGAGGCTGT TTTTTCTTTC CCTGAAGAAG ATAAAATTGC 2760

AAGCTCTTGC TCATTTGGAC GCTTAATGAA TTGTGAAGCT AGGAGCATGA TAATAAAGTA 2820

ACTTGCTCCT AAAATATAAA AAGTTTCTAC AAGCCCTACC CCTGCGATGA GGTGTTGCGC 2880

TATGGGACTA GTCAATAAAG AAGCAAAACC AAACCCCATA ATCGCTAAAC CTGTTGCGAG 2940

ACCACGTTTA TCAGGAAACC ATTTTATAAT CGTCGACACA GGGGTAATAT AGCCTGCTCC 3000

CAAACCAAGC CCACCTAAAA TGCCATAAGC GAGATACAAC AACCACAGCT CTGACGGTCT 3060

ATTGCAAATC CTGTTAAGAT ATTTCCACCT GCGTATAGAA AAGCAGATAG ACTTCCCATG 3120

ACTTTCGGAC CAAATTTTTC TACCAAACGC CCCATAAATG CAGCCGATAA GCCCAAACAA 3180

AAGATTGCTA GACTAAAGGC GAAGGCAACA GAAGCCTGAT CCCATCCCGT 3230

(2) INFORMATION FOR SEQ ID NO: 204:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5096 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

CCTATGAAGA CTGTCCCAAC TGGGTGTCCT TCTAGGCTAT CTGGTCCTGC CACTCCAGTC 60

AAACTAATTC CAAAATCAGA CTGGGTCTTG CTTCGTGCCT GCTCAGCCAT CTTCTGAGCT 120

GTAAATTCAG ACACCACACC ATGTTCTTCC AAATTCTTGG CAGGAATATC CAACATCCTT 180

GATTTTTCCT CCAAGCTATA GGTCACAAAA CCACCCTTAA ATATACTTGA AACTCCAGAA 240

AAATTCGCCA CGGTAGCTTG GAAAAGACCT GCCGTCAAAC TCTCTGCAGC CGCGATGGTT 300

TTCCCTTGCC TTTTCAGTTC TTCTACCACA ATGCTGGCTA AACTAGTTTC TTCCCCATAA 360

CCATAGCAAA AGTCTCGTAA AGAAATTCCT TCGAAAGTCT GGCAGTCCAA GATTTGATTT 420

TCCAAGATAT CCAGCGCTTG ATTCGCCTCT TCTTGACTGC TAGCCTTTGT TGACAGACGT 480

AGAGTGACTT CTCCTGTCTT GGCATAAGGG GCCAAGGTAG GATCGATCTG ATTATCAATT 540

AAATCAGCCA AAATCGTAAC CAACTGGCTC TCGCCAATCC CAAAGAAACG AAGAACTCGG 600

GAATACAGCT TGCTCCCTGT CATCAACTTG GGTAGAAGTT GGTTTAAGAC CATGGGTTTC 660

AATTCACTTG GCGGACCTGG AAGGACGACA TAGGTCACTC CGTCTACTTC TAATTTTCCT 720

CCAACAGCCA GTCCTGTTTC GTTTGGCAGT GGAATCGCTC CTTCTACAAT TTGAGCTTGT 780

CTTTCGTTAT TCGGTGTTCG GGCATAGTCT GGTCGCAGGG TAAAAAAGAT ATCCAACTTC 840

TCCTGAGCCT GAGGATCAAA GACTAATGCT TTCCCTAAAA ATTTAGCTAG GGTTTGTTTG 900

GTTAGGTCGT CCTCAGTTGG CCCCAAACCG CCTGTCAAAA TCACCAGACT GCTACGTTGA 960

CTGGCAATCT CAAGCAAAGA CAAGAGACGA ACTTCATTGT CTCCTACAGC CGTCTGAAAA 1020

TATACATCTA CCCCAATCTC AGCTAGTTTT TCCGACAAAA ACTGGGCATT GGTGTTGACA 1080

ATCTGCCCTG TCAAAATCTC TGTTCCAACA GCAATGATTT CTGCTTTCAT GTTTCCTCCT 1140

ACCTATCTAT TCGTATTTTT TTGAAAAAAT CGCAGGAATT TTCCTACGAT TGATTTTTTT 1200

ATTTGTATCA AAAGTTAATT ATCTTCATCA CCAACAGGTG CTCTGCCAAA TAAATCTTCA 1260

AATAAAACCG CATTGGTTTC AAGCTGAGTA ACTTCTTCTT GTCCCAAAGA ACGTCGGAGT 1320

AGATTTTGCA TTTCCAACAT ATGTGCTCTC GAAACAATCT GGTAAGAAAC ACCTTGAAGT 1380

ATCTCTCCTT CACCCTGCAA CTGCTGAGTT TCAATGGTTT TAAATGAATC TTTATAGCCT 1440

AGCAAGTTAG GGATACTTTT TGCAGACAAA TCAATATTGG TCTGCATATT GTCACTCAAA 1500

GCTTTTAGAA TCTCTTGATA ATGACCAATG CTATTTAAAC TGAGAGCTTT TTCCATGACT 1560

TTTTGAATAA CTTCACGTTG ACGTTTTTGA CGACCATAAT CCCCCTCAGG ATCTTGGTAA 1620

CGCATTCGTG CATAGACTAG GGCTTCTTCT CCCCCAATAT GTTGCTCCCC AACACCGATA 1680

GAAATAGTAT TAAATTCTTC TTGGTCACTG ATAGAAATTG GGAAACCTAG GATATTATTG 1740

ACTGTAATAC CTCCTACTGC ATCCACTAGT TTTTGCAATC CTCTCATATT GACCATCACA 1800

TAGCGATCAA TATGGATATT CATCATTTTT TGAATGGTTT CTATAGCAAG CTCTGCTCCA 1860

CCATCTGCAT ATGCTGAGTT CAGTTTCGCT TCATGAGCCT GACCATTCCC TGATTCAATG 1920

CGCGTCAGAA TATCCCGCTC TAAACTCATC ATTGTTGTTT TTTTCGTTTT AGGATTCACT 1980

GTCATCAAGA TCATGCTATC ACTTCTACCG ACCCAAGTTT CAGTTCGTTC AACATTTCCG 2040

GTGTCCACTC CCATTAACAG AATGGTTAGA GGTTCAGTCG CTTCAATAAC CTTGGTTTCT 2100

TCACCGATTT TTTTATAGGT TTTAGCTAAG GTTTCTGTCC CTTGTTGATA AATAGTATAA 2160

GCAAAAACAC CTACTCCTAC TACAGTTACA GAAAGTAAAG CTAGCACCAT TCCAATAATT 2220

TTTTTAACCA TATTTCTACT AACCTATCAG TTTACCCATC AAGTAAACAT CGATAAATTT 2280

CCCTTCTTCT ATATATGCCC CACGCTCTTG GCTACCTTCA ATGACAAAGC CATGCTTTTG 2340

ATAAAGATGG ACTGCTGCTT GATTACGAGT TTGGACAGTC AGTTGGAGAC GACGCAGAAT 2400

GCCACTTGCT TGTGCCCACT CTATCGCTTC TTCTAGCAAC AAACTTCCCA AGCCATTATT 2460

CCAATATCTT TTTCCAATCA CAATGAAGAG ATCTCCAATA TGACGGACTC TCTTACGCTG 2520

ATCAGCTGTA ATATTTACAA TACCAGCAAT TTTGCCATTT AAGAATGCAA GTAAGGTTAT 2580

CTGATTGTCC GAACTAGCTT GCTTGTTGAG GAATATTTCC ATCTCCTCAC TAGTCAAGAG 2640

AATACCATCT CCGTCTAGGC TGGTAAAGTC TGTCTCCAAA CTCACACGAT TTAAAAAGGC 2700

CACTAATTCA GCTGCATCTT TGGGCTCTGC TTCCCTAATG AGCAATTCAT ACTCCATATT 2760

GAAGCTCCTC TAACAATTTC TCAGCACGCA AACCCTTTGC CTGAAAATTT AAACGGCGTC 2820

CATCTGCTTC TTTTAGAATT TCCAATTCTA AATAAGCATC TGGCAAGGCA TCTCCTAAGA 2880

GATTTCCCCA CTCAATAACA GTCACGCCGC CACCAAAGAT AAACTCATCC AAGTCGATAG 2940

AATCAGCATC TCCTTCAATA CGATAAACAT CTAGGTGATA AAGTGGAAGT CGACCTTCAT 3000

ACTCTCTCAC GATAGTATAG GTGGGACTTT TAATCATTTG AGAAATCTGT AATCCTTTTG 3060

CAAGTCCTTT AGTAAAGGTC GTTTTACCTG CACCCAGTTC TCCAGTTAAG ATTAAAACAT 3120

CATTCTTTGC TAATAGATGG CCCAAACGCT CCCCTAAGGC TTGCAACTCT TCTTCATTTT 3180

TTGTGTACAT ACTCTTATTA TACCAAAAAC TTTTCTTTTG TGTCTATTTT CCTACTAAAC 3240

TTATCATCAT AACATCCATA AAAAACAGGC TTTCTCTAAA AGAAAATGAG CGTAACAATG 3300

ACCAATACAA GATCTCGGAA AATATGACCA TAAAAGGAAA CTTCCTTCTT AACCGAATTT 3360

GGGACAAGAT AGGCTGCAAA AAACAAGCCC AGTCCAATAT AAATCAGAAG TGAGACAATG 3420

GTCATTGGAT TTCTTAAGAA AAGAAGTGTT GCTAAAATAG TCACCAACAC TGTCTTTTTT 3480

CTGTCCAGCA TAGCAAGAAA ATCGCGCACG TATTTTTTCA AGGGTAAAAA AATCAGCAAA 3540

TCTAGCCCAA ATAGGAAAAA GAAGGATGGC AATAAAAAGT CAACTAATTC TTGCTGCAGC 3600

GTATTTTTGA TGAACAAGTT ATCTGACAAA ACAAGAACAG CTCCTAACAA ATTAATTAAG 3660

AGTAACATAC TGTAAAAAAG CTTCACCGAC TTCTTACTGG CTAGGACACT ATGGACTTCT 3720

TGCTTACGGG TATAAAGATA ATTTACTCCA GCACAGATTC CTGAAACGAA AACCATGCTT 3780

CCGATGAAAA AAGCTGTACT TTGTTTAAAG GACAAGATGC ATTCCTTCCA TAGGAAACAG 3840

CTACTCAAAC TGATTTGAAT TAAAGCTAAC AAAAATAAGA TTCTCATTGA TTTCATCTTC 3900

TCTCTCCCTT CCTACCAATC ATTATACTAG GAGAAAAGAG AGAACTGTTT CTAATCTTCT 3960

CAAATGTCTC TTTAAGACGC TAAACAAACA CTAGAGACTA ATACTCAATG AAAATCAAAG 4020

ATCAAACTAG GTAGCTAGCC ACAGGTTGCT CAAAACAGTG TTTTGAGATT GCAGATAGAG 4080

CTGACGTGAT TTGAAGAGAT TTTCGAAGAA TATAAATTTG AAATCATGAA AATCCGTCAA 4140

ACGGGTGGTT GTTTTGTCTC GCACCTCACG GAGCGAGACG GACTCAGAGT CACATAATTA 4200

TAAGGCTGAT AGTATTAATC TAACTATCAG CtTmCAGGTT ATTTAACGTT TCAGAAAAAC 4260

TATAATGTCA AGATTAACTA AACAGTATCT AGTTCCTTCA AATAATTTTC TATCTTCATC 4320

AACATTAAAG GATTGTTATA AATCTTACAT AACTCTCTTG CTTCTATATA ATAATTTTTG 4380

ACTTGTTCTC TGTCTAGAAA TTTGGCTCCA GCATTTCCTA CAAGAATAAG TAGAGGAGCC 4440

AATTGGTAGC TTGTCTGTCT TTGTTTACAG AGTTCAATCG TTTCAAGAGC TTCTTGGATG 4500

GCTTCATTAT ATTTTTCCTT TGATACTAGG TAGTGAGCGT AGTTGTAACG AACTCTGATG 4560

TAGCCAAATA AAAACTCTTG ATGGTCCAAA TTTTTTGTCT GATACAACTC TATTAAATGA 4620

GAGTAGTTTG CCTCATATTC TTGTTCACGA CCCACTAAGG AATAGAAATT AGATAGAGTA 4680

TTCAACGCCT TTAAATAAAT CAGAGTATTT GAAGAGACTT TTAATAATAT ATTTTCCAAT 4740

GACGAAATTG CCTCACACTT ACTGTCATAT TGATAGAAGT CAATTATAGA TTTAATCCAT 4800

TCAAGGTAAG TTCGGTCTTC TAATGTTAGA AAAGTGCTTC GTTCTACTTC TATTTTATAA 4860

AGATATTCTA AATCGTCATA ATTTCTGTCA TCTAATAGGC GAGCAGATAG ATGTTTGAAA 4920

TTAGAGAGGT TAGACTTAAC TTCGATTTGT TCATTGAAAA AGTAATCCAA AGGGACTTCA 4980

AGTCGTTGAG AGAGTTTGAA TAACAAGTCT GCGGAGGGAA TAAAATGACC TCTTTCAATT 5040

TTACTAATCT GGCTTTGTTC ACAAATTCCT TCTGCAAGAG TTTGTTGGGA GAGTCT 5096

(2) INFORMATION FOR SEQ ID NO: 205:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2395 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

ACAAGATAAA AATAAAGGAT TACAATGGGG AATATAAAGT AAACCGGTAA ACCTAAAAAG 60

AAAGGAGAAA AGATGAAAAT TGTACTTGTA GGGCATGGAC ATTTTGCTAC AGGGATTTAT 120

AGTTCTTTAC AATTGATTGC AGGTAATCAA GAAAATGTGG AGGCGATTGA CTTTGTGGAA 180

GGAATGTCAG CAGATGAACT CAAGCAAAAA ATCTTACTTG CAATTTCAAA TGAAGAAGAA 240

GTTTTAATCC TAAGTGATCT CTTGGGAGGA TCGCCATTCA AGGTTTCTTC TACCATAATG 300

GGAGAAAATC CAGCCAAGAC AATGAATGTT CTCTCGGGTT TGAACTTAGC CATGTTAATG 360

GAAGCAGTCT TTGCTAGAAT GGCTCATAGC TTTGATGAGG TTGTTAATAA ATCAGTAGTG 420

GCGGCCCAGG GCGGAGTCGT AAATGGTAAA GAATTGTTTT CAACGGATGC AGAGGAAGAG 480

GAAGAAGATT TCGAATCGGG TATTTAAAGG GTAAAAGAAT GATAAAAAAG GTTACGATTG 540

AAAAAATAAA ATCGCCTGAG CGCTTCTTAG AAGTACCACT TCTGACGAAA GAAGAAGTCG 600

GCCAGGCAAT CGATAAGGTT ATTCGGCAGT TAGAACTCAA CCTTGACTAT TTCAAGGAAG 660

ATTTCCCGAC GCCAGCTACC TTTGATAATG TCTATCCAAT CATGGATAAC ACGGAATGGA 720

CCAATGGTTT CTGGACAGGA GAACTGTGGT TGGCTTATGA ATACAGTCAA CAGGATGCAT 780

TTAAAAACAT CGCTCATAAA AATGTTCTTT CTTTCCTGGA TCGTGTCAAT AAGAGAGTAG 840

AATTGGATCA CCATGATCTC GGCTTCTTGT ACACACCGTC TTGTATGGCT GAATATAAGA 900

TAAATGGAGA TGGAGAGGCT AGAGAAGCAA CCTTGAAAGC TGCAGATAAG TTGATTGAAC 960

GCTATCAAGA AAAAGGTGGT TTTATTCAAG CTTGGGGAGA CTTGGGCAAG AAAGAGCATT 1020

ACCGTTTGAT TATCGACTGC TTGCTCAATA TCCAACTCTT ATTCTTTGCT TATCAAGAAA 1080

CAGGCGATCA AAAATACTAC GATATTGCAG AAAGCCATTT CTATGCTTCA GCTAATAATG 1140

TAATCCGTGA TGACGCTTCG TCCTTCCACA CCTTCTATTT TGATCCTGAG ACAGGTCAAC 1200

CCTTTAAAGG TGTAACGAGA CAAGGGTATA GTGATGATTC ATGCTGGGCA CGTGGTCAAT 1260

cATGGGGAGT CTATGGTATT CCTTTGACTT ATCGTCACTT AAAAGACGAG tCCTGCTTTG 1320

ACTTGTTTAA GGGTGTGACC AATTATTTCT TGAATCGTCT GCCAAAAGAT CATGTGTCCT 1380

ATTGGGATTT GATTTTTAAT GATGGTAGTG ATCAATCACG AGATTCTTCA GCAACAGCTA 1440

TCGCCGTCTG TGGGATTCAT GAAATGCTAA AACATCTCCC AGAGGTGGAT GCTGACAAAG 1500

ATATTTATAA ACATGCTATG CATGCCATGC TTCGTTCCTT GATCGAACAT TATGCAAATG 1560

ATCAATTTAC CCCTGGTGGG ACAAGTCTCC TCCACGGTGT GTACTCATGG CATTCAGGTA 1620

AAGGAGTGGA TGAAGGCAAT ATCTGGGGTG ACTACTATTA CCTAGAAGCC CTTATCCGTT 1680

TCTACAAAGA CTGGAACCTA TATTGGTAGG AGGAGAAATA TGACAATGCC AAATATTATT 1740

ATGACCCGTA TCGATGAACG GTTGATTCAT GGACAAGGAC AACTTTGGGT AAAATACCTA 1800

GGTTGTAATA CGGTCATTGT TGCCAATGAC GAAGTAAGCA CGGACAAGAT GCAACAAACT 1860

CTGATGAAAA CAGTTGTGCC AGACTCAGTT GCCATGCGTT TCTTCCCTTT GCAAAAGGTG 1920

ATTGATATCA TTCACAAGGC TAATCCTGCT CAAACGATCT TTATCGTTGT AAAGGATGTG 1980

AAGGACGCTT TAACCTTGGT AGAAGGTGGT GTCACTATCA AAGAAATCAA TATTGGGAAC 2040

ATTCACAATG CCCCTGGTAA AGAGCAAGTG ACACGCTCCA TCTTCCTGGG TGAAGAGGAC 2100

AAGGCGGCCC TCAAGGAATT GAGCCAAACT CATCAAGTAA CATTTAATAC GAAAACAACT 2160

CCAACAGGAA ATGATGGAGC TGTTCAAGTC AACATTATGG ACTATATTTA ACAGAGGAGA 2220

TCGTTATGTC GATTAATGTA TTTCAAGCGA TTTTAATTGG ATTATGGACA GCTTTCTGTT 2280

TTAGTGGAAT GCTGTTAGGA ATTTACACCA ATAGATGTAT TGTTCTGTCA TTTGGTGTCG 2340

GAATTATTCT AGGTGATCTG TCATGCTCTT GCAATGGGAG CCAATGGTGA ATTGG 2395

(2) INFORMATION FOR SEQ ID NO: 206:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3342 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

CCTTCTTTAG AGGTTAATTT TGCAAAATCG TCGATTGTTA TATAAGGATT ATTATAGAGA 60

CTGTTCGCAA AGAATCTCTG ATATGTTTTT GAATCTTTTG AATACAAAAC TATCTCTCTA 120

ATAGCATTGC CATCTGTTCC ATCAATTGGT AAACATACCG TAACTAGAAA AAGAATTATA 180

TTCAAAATAA AAAATTCTGA TGCGTACGGC ACAAATCCCA AAAGTGCTAA TATTGCGACA 240

ATTAGGTTAG CTCCACCTCC CCCAAAGAAG TAGAACACCA AATTCCTATC ACTATTTTTT 300

TCATTAGTAA TGTTTCTATT ACTCATTTGA CAATAACCGA ATGCTAATAA CACTGGAAAT 360

TTGAAATATA TTTTTTTTCT GAAATAGAAG AAAAAGGGAG TAGCAAGCAT CTCTAGTTTA 420

TAAGATAAAC ATCTTCCCAC TAAAAAATGA CCTAGTTCAT GTAATGTAAT TGATATTAAC 480

GAAATTAAAA TCAATCGAAA ATAATAGATT AATGAATCAT TTGGAAAAAT TATCAATAAT 540

AGGAACAATA ACGGAATCAA ACATAAATAT ATGACAGAGT TATTTAATAT TTTCAACATA 600

ATACCATTCC TCTAAACTAT TAGCTTCAAA AAGGCGTTTT TTCTCCCAAT ACATCTTCTC 660

AAAATGTTCG GAATCATAAT TTTCTAAAAT TAATTTTAtG TCTGGTAAGC TCTTTCTTGA 720

TAATCCGTTG TTTTGTACTT AATTTTCCCT TCAAGTACAT CTTCAATTTT ATAAGTTGCC 780

TCCATCAACT GAGCCTCTGC AATATCTTTG AGTGAATTGG TAATTGAAAC TTGGTGTAAT 840

ATCTGTCCts CCATATATGA AAATATATCT CTAAGATATT CTGACACATT ATCAGAGCCG 900

TTACTCTCAG CAACATCTAA TGTTACAACA AACTTTCCAG CTAATCGAAA AAGATGGCTC 960

CACCCCCCAA TCCTTTCAAT AAAGTTTTTT GTGTCCACAG ATACGTTTTG TAAATATACA 1020

GGAGAAGAGA TAATTATAAT ATCAGACTCT AATAACTCTT TTTTTATAAC ACCTCCATCA 1080

TCAGCATTAC TTTGCCTATC AATTCCTTTC TTAAACAACT CTTCTGAATC AGAATTAGAT 1140

ATTTCTAGCT CTGAATTGAA AGGTGTCCTG AAAGATATAT CAACATTATT TCTACTAGAA 1200

ATGATACTTG AAAGTCTCTT AGTATACTCT AAAGTCTTAG AGTTATGATT TCGCACTCCT 1260

GCATATATAA ATATTTTATT CATTTTAATT CATCCTCTCA ATTTGAATTT AGTAGATTTT 1320

TCAAGATAGT ATGGTACAAA AACAGACTTT TGTTGACTCA CATTATTACA TATGTTTTGT 1380

ATTAAACCAA AATCAATACT ATTTTTGGAG TAATTTTGAT TTTAGTTTAA AATCATTTCT 1440

ATAACAGTAG CATATACCTC AAGCCGTTTA GCAATTAGAA TAGAACTTTT CTTTATTATA 1500

TTATTATCTC AACGAAAAGC TACACTATTA AAAATATTTT ATAGAATTAC ATATTAAACT 1560

AGTCAATCTT GGTATTTTTA TATTGCTTAA TGAGTGGACA CCTCTATTTT AGAAACAAAA 1620

CTATAAATTA AGCTAGATTT CAAGTAATGA GGGGATAACT ATCTTTTTGT CATTCTGATT 1680

CAGTGCGATA TACCTTAAAA AAGTATAAGC AATACCAGTC ACACCTGTAT ACAAAGAAAA 1740

ATCTGGGAAA TTGCTTGTTT GGACGATACG ATACTCTCCT TCTTTTGATT TATTCATTAC 1800

AACACTACAC AATAAAGACT CCAATTCCAT ACTAGTATCC ATTTCTTTCA TGTAGTCGAT 1860

GTAAAAATTT ATTATGGCCA TACTTCCATG GCAAAATGTA TCATTATCTA AACTAGCTAC 1920

AATTCCCTCT GGAACACTTT GGGGATGATT AACTAATGTC CCAAATTCTC CACTACACCA 1980

CTTCAAAGAA TGAATTTTGA TTTTCTCCCT AGGAACTAGT TGTAAAATTA ATTCTTTATA 2040

TTTTTTAAGT CTTGTCACTT TATAAATATT TTTTAATGTA AAAATTACAC CTGATAGTCC 2100

ATGGCCAAAA CTATATCCAA AATTACTATT ATCTCTCTCG CTTACATCAT TATATAGCGT 2160

ATCACCTAAA CTTAATACTA GCCTTAGAAC ACGTTCCTTC TCTATTCCTC TCCTATAATA 2220

TCTTACCAGT GTATTAATTA AAGGTAGAAG ACCATTAATA TAGTCAGACT TGTTTGAAAC 2280

ACTTGCAAAA TCAGTCTTTT CAAGCTCAGT TAAAACACTC TTTATATAAT TTAAGCATGC 2340

GAGAGTATTT GTATCGTAAT CCTCTATAAT GGATAGAACA ATGAAATATC CTATATCCCC 2400

AGTTAAACCA AATGTGGTCT TAGATAAAGA AACAGATGGC GGAATTGCAG ATAACATTTT 2460

ATTGTACAGT TGAGTATATG ATGATTTATC TTTCAATAAT TTTACATAGT ACATAAACAG 2520

TAATATTCCA GCTCTACCCC TATACATATC ATTmCCCGTT TGTTCAAGAC ACCATTTAGA 2580

ACCTTTAAAA TTAACAGGTA TACTCCAAAT TGGATATTCG TCATAAATAT TATTAATAAC 2640

CAAAGAGTCT GCAATATTTT CTACTTCATT ATGCAGAATA GTAACTAAAC TTTCATTTGG 2700

GAGTTTTTTT CTATTAGATA AGTTTAATTT ATATCCTTTT TTTCGCTGAT CAAAGCTTGG 2760

AAAATAAATT TCAATGATAT CAAGTTGCTT TTCTAAATTT TCCAAATTAT TATTAGGTAA 2820

ATATTTCATA AAATAGTCAT ATCCAGAAAA TTGATGTAGG GAAATAAAAT GATTTCCAAA 2880

ATCATCGTAG ATTTCATTGA TATTTGTATC TGTATAAAAA ATCGGAATAT CTAATAACCT 2940

CATTTGTTCA CATTCGCTTG CTACAATACC TTGATTAGAA AACTTATTGC TCCAGAGATT 3000

TTCCAATGCT TTTTCTCTAT CTAACATTTC TTCATAAAAA TCAGGATGAT ATAAAAAAGA 3060

TAGTACTGAA GCATAGCTAT TTGTGTCTCT AAAAAGTACC CTTGTCTTTA AACCATACAA 3120

GTTTGCTTTT AATAGCATTT TAAATTCTTC TGTTTTATTT AACTCTTCAA ATATCAGATA 3180

AAAATCCCTA AAACCTTTTT TGAAATCTTT TATATACTTA TCAAATTCTA TATCACCATC 3240

CCGAACAGGC AGGTTTTTCC CACCTTCAAA ATCAATTTTC CCAATATCAA ACTTTACCTT 3300

ATCAGTATTT AAATTAATTA AAACTTGACC AGGGATCCTC TA 3342

(2) INFORMATION FOR SEQ ID NO: 207:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3454 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

GAGAAAAGAA TGTTAAAGAA AAATGATATT GTAGAAGTTG AAATTGTTGA TTTGACCCAT 60

GAAGGGGCAG GAGTTGCCAA GGTAGATGGT TTGGTCTTTT TTGTAGAGAA TGCTTTACCG 120

AGTGAAAAAA TTCTCATGCG TGTCCTCAAG GTCAATAAAA AGATTGGCTT TGGAAAAGTT 180

GAAAAATACC TTGTCCAGTC ACCACACCGT AATCAAGATC TAGATTTGGC TTACCTGCGT 240

TCAGGAATCG CGGATTTAGG ACACCTTTCT TATCCAGAAC AGCTCAAGTT TAAAACCAAG 300

CAAGTCAAGG ACAGTCTCTA CAAGATTGCT GGAATTGCAG ATGTAGAAGT TGCTGAAACG 360

CTTGGTATGG AACATCCAGT CAAGTATCGC AATAAGGCGC AGGTGCCCGT TCGTCGAGTG 420

AATGGTGTCT TGGAAACAGG ATTTTTCCGT AAGAATTCGC ATAACCTCAT GCCCCTTGAA 480

GATTTCTTTA TCCAGGATCC TGTCATTGAC CAAGTCGTAG TAGCTCTTCG AGACCTGCTC 540

CGTCGTTTTG ATTTAAAACC TTATGACGAA AAGGAACAGT CTGGATTGAT TCGGAATCTT 600

GTGGTGCGTC GTGGTCACTA TTCAGGACAA ATCATGGTCG TTTTGGTGAC AACTCGTCCA 660

AAAGTTTTTC GTGTTGACCA ATTGATTGAA CAAGTTATCA AGCAGTTCCC AGAGATTGTG 720

TCTGTCATGC AAAATATCAA CGACCAGAAT ACCAATGCGA TTTTTGGTAA GGAGTGGCGC 780

ACTCTTTATG GTCAAGACTA TATTACGGAC CAGATGTTGG GAAATGACTT CCAAATCGCT 840

GGCCCAGCCT TTTACCAAGT CAATACTGAA ATGGCGGAGA AACTCTATCA AACAGCCATT 900

GACTTTGCAG AGTTAAAAAA AGATGATGTG ATTATTGATG CCTATTCTGG TATTGGAACC 960

ATTGGTTTAT CAGTCGCCAA GCATGTCAAA GAAGTCTACG GTGTTGAACT GATTCCAGAA 1020

GCAGTAGAGA ATAGCCAGAA GAATGCTTCT TTGAACAAGA TTACTAATGC CCACTATGTC 1080

TGTGACACGG CTGAAAATGC CATGAAGAAA TGGCTCAAGG AAGGTATTCA ACCAACCGTT 1140

ATCTTGGTTG ATCCTCCACG CAAGGGCTTG ACAGAAAGCT TTATCAAAGC AAGCGCCCAA 1200

ACAGGAGCCG ATCGCATCGC CTATATCTCC TGCAATGTCG CAACCATGGC GCGTGATATT 1260

AAACTATACC AAGAGTTGGG ATATGAATTG AAGAAAGTCC AGCCGGTGGA TCTATTTCCT 1320

CAAACGCATC ACGTCGAGAC GGTAGCACTT TTGTCCAAAC TCGATGTCGA TAAGCACATA 1380

AGTGTTGAAA TTGAGCTGGA TGAGATGGAT TTGACAAGTG CGGAGAGCAA AGCAACATAT 1440

GCTCAAATCA AAGAATATGT TTGGAATAAA TTTGAATTAA AAGTTTCGAC ATTATATATT 1500

GCACAGATAA AAAAGAAATG TGGAATAGAA TTACGAGAAC ATTACAACAA GTCTAAAAAG 1560

GATAAACAAA TTATTCCACA GTGTACACCT GAAAAAGAAG AAGCCATCAT GGATGCTTTG 1620

AGACACTTCA AAATGATTTA ATAGAAAAGA ATGACAGTAT ATGACTTTCT GCATTTATTA 1680

CATTCCTACT TGGTATAGGA ACAGCTATTA TTCCTTTCTT GCAAGGTATC AATTAGAAAA 1740

TAGGCTCAAT ATAAAGATTG ATAGGATCAT TTTTATATTT AAAGGAGCGT TGAAATGATT 1800

GATAAAGGCA ACAAAAAATT TTAGGATAAA TTTGCTAAGT TGTATGCCTC TTTTATGAAA 1860

AAAGATAAAG AGGTTTATGA TAAAGTTTGT GAATATCTTA GTCCTCATTT GAATAAAGAT 1920

ATGGAGGTGC TTGAACTTGC TTGTTGGTTT CGTGTCATAA CAGTTATAGA GGCAAATAGT 1980

TATGTAAATA TAAGGAGTTC AAGACTTCTA CCAAAGTTTA AAACTCAAAA AATAAATAGT 2040

TGGTGTGCTG CTTACAATAT CCATTTTAAT AATGGATATT GTAAGCAGCA CCCCcAtGAA 2100

TTTAAAGATT CTTTAAAGAG TCTTATTTTG TGATGAAAAT TTAATATGTA AATCTCAGAC 2160

GATAGAAATT AAAAACTCTA TCGTCTTTTT TATACTCAAA ATTAGGAGGT AAAAATGGTA 2220

AGGATAAGAG GTCCCACTTA AAACAATTTA TGGCAAAATA AGGACGGAAT AACACAACAA 2280

ATTCTCTAAA ACAAATCACT AAATCAATGT AAGATTGAAT GAAATCAATA TTTATGCTAT 2340

AATTAAATAA ATTTAATGAA GAAAAAAAGA GGGATATTAT GGCACTTAAC TATAAACCAT 2400

TATGGATACA GTTAGCAAAA AAAGGACTAA AGAAAACAGA TGTAATAGCT ATGGCAGGAC 2460

TTACAACAAA TGTTATGGCA CAAATGGGAA AGGATAAACC AATTACATTT AAGAATTTAG 2520

AAAGAATATG TAAGGCTTTA TCTTGCACTC CTAATGATAT TATTAGTTTT GAAGATAATT 2580

TTAGTGACGA GGAATAGAAA ATGACTTTAA GGACAGAAGA TCAAGTTAGG GATTATGCAA 2640

GAGAAGTATA GGCTTTAATG AAGTTGAAGA AAACATCAAT CAAGGTACTG GTCAAATAAC 2700

TACTTTTAAT CAATTAGGCT TCAAGGGATA TTCAAATAAG CCAGATGGTT GGTATTTACC 2760

TAAAAATATG AATGATGTAG CAATAATCCT TGAAACAAAA TCAGAAGAAA GAGATATTAG 2820

CAAACAAATT TTTATTGATG AGTTAATGAA AAATATAGAC ATAATTTAAC TAAAAATAAA 2880

AACTAGATCC TTTTTTGAAA AAATTATATT ATTAAATTTG TAACTGTATC TATTGACAAT 2940

GATAATTATT ATCGATACAA TAGACTTGAA ATATGTTTAA GGAGTTTTTA TGAAAaCAAA 3000

TTTTTTCTAA TmGCTATTTT AGCTATGTGT ATAGTTTTTA GCGCTTGTTC TTCTAATTCT 3060

GTTAAAAATG AAGAAAATAC TTCTAAAGAG CATGCGCCTG ATAAAATAGT TTTAGATCAT 3120

GCTTTCGGTC AAACTATATT AGATAAAAAA CCTGAAAGAG TTGCAACTAT TGCTTGGGGA 3180

AATCATGATG TAGCATTAGC TTTAGGAATA GTTCCTGTTG GATTTTCAAA AGCAAATTAC 3240

GGTGTAAGTG CTGATAAAGG AGTTTTACCA TGGACAGAAG AAAAAATCAA AGAACTAAAT 3300

GGTAAAGCTA ACCTATTTGA CGATTTGGAT GGACTTAACT TTGAAGCAAT ATCAAATTCT 3360

AAACCAGATG TTATCTTAGC AGGTTATTCT GGTATAACTA AAGAAGATTA TGACACTCTA 3420

TCAAAAATTG CTCCTGTAGC AGCATACAAA TCTG 3454

(2) INFORMATION FOR SEQ ID NO: 208:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3752 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

CGGGAGTATA CTTAATATAA TTATAGTCTA AAAATGACTA TCAGAAAAGA GGTAAATTTA 60

GATGAATAAG AAAAAAATGA TTTTAACAAG TCTAGCCAGC GTCGCTATCT TAGGGGCTGG 120

TTTTGTTACG TCTCAGCCTA CTTTTGTAAG AGCAGAAGAA TCTCCACAAG TTGTCGAAAA 180

ATCTTCATTA GAGAAGAAAT ATGAGGAAGC AAAAGCAAAA GCTGATACTG CCAAGAAAGA 240

TTACGAAACG GCTAAAAAGA AAGCAGAAGA CGCTCAGAAA AAGTATGAAG ATGATCAGAA 300

GAGAACTGAG GAGAAAGCTC GAAAAGAAGC AGAAGCATCT CAAAAATTGA ATGATGTGGC 360

GCTTGTTGTT CAAAATGCAT ATAAAGAGTA CCGAGAAGTT CAAAATCAAC GTAGTAAATA 420

TAAATCTGAC GCTGAATATC AGAAAAAATT AACAGAGGTC GACTCTAAAA TAGAGAAGGC 480

TAGGAAAGAG CAACAGGACT TGCAAAATAA ATTTAATGAA GTAAGAGCAG TTGTAGTTCC 540

TGAACCAAAT GCGTTGGCTG AGACTAAGAA AAAAGCAGAA GAAGCTAAAG CAGAAGAAAA 600

AGTAGCTAAG AGAAAATATG ATTATGCAAC TCTAAAGGTA GCACTAGCGA AGAAAGAAGT 660

AGAGGCTAAG GAACTTGAAA TTGAAAAACT TCAATATGAA ATTTCTACTT TGGAACAAGA 720

AGTTGCTACT GCTCAACATC AAGTAGATAA TTTGAAAAAA CTTCTTGCTG GTGCGGATCC 780

TGATGATGGC ACAGAAGTTA TAGAAGCTAA ATTAAAAAAA GGAGAAGCTG AGCTAAACGC 840

TAAACAAGCT GAGTTAGCAA AAAAACAAAC AGAACTTGAA AAACTTCTTG ACAGCCTTGA 900

TCCTGAAGGT AAGACTCAGG ATGAATTAGA TAAAGAAGCA GAAGAAGCTG AGTTGGATAA 960

AAAAGCTGAT GAACTTCAAA ATAAAGTTGC TGATTTAGAA AAAGAAATTA GTAACCTTGA 1020

AATATTACTT GGAGGGGCTG ATCCTGAAGA TGATACTGCT GCTCTTCAAA ATAAATTAGC 1080

TGCTAAAAAA GCTGAGTTAG CAAAAAAACA AACAGAACTT GAAAAACTTC TTGACAGCCT 1140

TGATCCTGAA GGTAAGACTC AGGATGAATT AGATAAAGAA GCAGAAGAAG CTGAGTTGGA 1200

TAAAAAAGCT GATGAACTTC AAAATAAAGT TGCTGATTTA GAAAAAGAAA TTAGTAACCT 1260

TGAAATATTA CTTGGAGGGG CTGATTCTGA AGATGATACT GCTGCTCTTC AAAATAAATT 1320

AGCTACTAAA AAAGCTGAAT TGGAAAAAAC TCAAAAAGAA TTAGATGCAG CTCTTAATGA 1380

GTTAGGCCCT GATGGAGATG AAGAAGAAAC TCCAGCGCCG GCTCCTCAAC CAGAGCAACC 1440

AGCTCCTGCA CCAAAACCAG AGCAACCAGC TCCAGCTCCA AAACCAGAGC AACCAGCTCC 1500

TGCACCAAAA CCAGAGCAAC CAGCTCCAGC TCCAAAACCA GAGCAACCAG CTCCAGCTCC 1560

AAAACCAGAG CAACCAGCTA AGCCGGAGAA ACCAGCTGAA GAGCCTACTC AACCAGAAAA 1620

ACCAGCCACT CCAAAAACAG GCTGGAAACA AGAAAACGGT ATGTGGTATT TCTACAATAC 1680

TGATGGTTCA ATGGCAATAG GTTGGCTCCA AAACAACGGT TCATGGTACT ACCTAAACGC 1740

TAACGGCGCT ATGGCAACAG GTTGGGTGAA AGATGGAGAT ACCTGGTACT ATCTTGAAGC 1800

ATCAGGTGCT ATGAAAGCAA GCCAATGGTT CAAAGTATCA GATAAATGGT ACTATGTCAA 1860

CAGCAATGGC GCTATGGCGA CAGGCTGGCT CCAATACAAT GGCTCATGGT ACTACCTCAA 1920

CGCTAATGGT GATATGGCGA CAGGATGGCT CCAATACAAC GGTTCATGGT ATTACCTCAA 1980

CGCTAATGGT GATATGGCGA CAGGATGGGC TAAAGTCAAC GGTTCATGGT ACTACCTAAA 2040

CGCTAACGGT GCTATGGCTA CAGGTTGGGC TAAAGTCAAC GGTTCATGGT ACTACCTAAA 2100

CGCTAACGGT TCAATGGCAA CAGGTTGGGT GAAAGATGGA GATACCTGGT ACTATCTTGA 2160

AGCATCAGGT GCTATGAAAG CAAGCCAATG GTTCAAAGTA TCAGATAAAT GGTACTATGT 2220

CAATGGCTTA GGTGCCCTTG CAGTCAACAC AACTGTAGAT GGCTATAAAG TCAATGCCAA 2280

TGGTGAATGG GTTTAAGCCG ATTAAATTAA ATCATGTTAA GAACATTTGA CATTTTAATT 2340

TTGAAACAAA GATAAGGTTC GATTGAATAG ATTTATGTTC GTATTCTTTA GGTACCTATC 2400

TTATGATTTC AGGAAATGTC ATTAAAAAAA CGACTCATTT TCTCTAACCT GAAAAATAGA 2460

TTAGAGAAAA TGGGTTGTTT TATCTATTAT AGTTATTTGA ATGAAGmTAA GAAGAAGGTA 2520

TACTCACATC ATTCACATAA TCTGTATATT GACTATAAGT TTTAAAAAAC AATTTTTAAG 2580

CTCTTCCTTG TCTTCTCTAA CCAAGCGTGT TATAATGAAT ACTGCTCAAG CGACCTTCAA 2640

TCGTGAAGCA CACACGACCT TCAATCGTGA ATAAACGAAT AGATGGGAGA CTTACCATGA 2700

GTGATAACTC TAAAACACGT GTTGTCGTGG GGATGAGTGG TGGTGTTGAT TCGTCGGTGA 2760

CGGCTCTTTT GCTCAAGGAG CAGGGCTACG ATGTGATCGG TATCTTCATG AAGAACTGGG 2820

ATGACACAGA TGAAAACGGC GTCTGTACGG CGACCGAAGA TTACAAGGAT GTGGTTGCGG 2880

TGGCAGACCA GATTGGCATT CCCTACTACT CTGTCAATTT TGAAAAAGAG TACTGGGACC 2940

GCGTTTTTGA GTATTTCCTA GCGGAATACC GTGCAGGGCG CACGCCAAAT CCGGACGTTA 3000

TGTGCAACAA GGAAATCAAG TTCAAGGCCT TTTTGGACTA TGCCATAACC TTGGGGGCAG 3060

ACTATGTAGC GACTGGGCAT TATGCTCGAG TGGCGCGTGA TGAGGATGGT ACCGTTCACA 3120

TGCTTCGTGG CGTGGACAAT GGCAAGGATC AGACCTATTT CCTCAGCCAA CTTTCGCAAG 3180

AACAACTTCA AAAAACCATG TTCCCACTAG GACATTTGGA AAAGCCTGAA GTACGCAGAC 3240

TAGCAGAAGA AGCAGGCCTT TCGACTGCTA AGAAGAAAGA CTCGACAGGG ATTTGCTTTA 3300

TCGGAGAAAA GAACTTTAAA AACTTTCTCA GCAACTACCT GCCAGCTCAG CCTGGTCGCA 3360

TGATGACTGT GGATGGTCGC GATATGGGCG AGCATGCAGG TCTTATGTAC TATACAATCG 3420

GTCAGCGTGG CGGACTCGGT ATCGGTGGGC AACACGGCGG TGACAATGCC CCTTGGTTCG 3480

TTGTCGGAAA AGATCTAAGC AAGAATATTC TCTATGTAGG ACAAGGATTC TACCATGATT 3540

CGCTCATGTC AACTAGCCTA GAAGCCAGTC AAGTCCACTT TACTCGTGAA ATGCCAGAAG 3600

AGTTTACGCT AGAATGTACG GCTAAATTCC GTTACCGTCA GCCTGACTCT AAGGTGACCG 3660

TTCATGTCAA AGGAGAAAAG ACAGAGGTCA TCTTTGCGGA ACCACAACGC GCGATTACAC 3720

CAGGACAGGC AGTTGTCTTT TACGATGGCG GG 3752

(2) INFORMATION FOR SEQ ID NO: 209:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3580 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

TATTTATATT TTTTTATCTC TGGCATACTT TGATACCTTT TTAGACTTAA AGTCTTTAAT 60

AGTGCCTTTC CACCTCTTTT TATCTATAAA GATTCTCCTA CATCATAATT CATTTTTTTA 120

TTTAAACCTT TCTGTCTTAG TTTGTCTTTA TCTTCTTCAT ACCATTTTAA GATTGTCACA 180

TAGTGGTTTT GATAGGTCTT ACCACTGCTT TCCATGTATC TGGATAGTTT ATTTATCATT 240

ATATCTGTGT GTGAGTTTAA TTTTTCTTTT AGATTTTTAT ATTCTTCTTT GCTTAACCTT 300

ACATTTTTGA ATTCTCCATA AAAAATGGGG GTGGACTTTT TATCTATCTC TCCCTCTCTC 360

TCTTTATCTA TCTCTATATC TTTCCATGTA ATTCCAATCT GGAGTACCTC TACTGTCTAT 420

CGGTAATTTA ATTTTGATAT CTGGCAATAC TGTGCTAGAT ATTTGATCTT TATATTCAGT 480

ATTTTTTAAA GCTTGCCTAA TAATTGAAGT TAAATAGAAT GCTACTTCTT TATTCAATTC 540

TTTATTTTTT AATTTTAAAC AATGAATTTT CATATCTAGG CTTGCTTTAT ATTTATGATA 600

AAAGACTGCT CCTAAAAATG AAACAGATAT AAAATTTTCA AAAACTCTAT AATTTTTATC 660

ATCTATATCT TCGTAGTAAC CTAAGATACC ATTGTCAATA TTTGTAGCAC TAATTCTAGG 720

AGTTTTTCCA TCGAGTAAAT ATCTTTTTGG AATAGATGAG CCTGTTGGTA CTTAACTCGA 780

TTTCCCCTTT TTTTCGGTAA TAAATATTTC TTTTTATTTT GTTGTCTGAT ATTTTTCCTA 840

CCTGTCCTTT GTAGGATGAG TATTTTCTAG ATTTTCyTGA ATAACTTTTT ACTTGAAGTT 900

TTAGCTTTTG AACTAGTCGT TGTACTTTCT TTTTGTTTAT TATCAGTCCT GATCTTTTTA 960

ATATTGCTGT TATTCTCTAT ATCCTATTTT TCATTCATGA TATTCTTTTA CTAATTTTAT 1020

CTTAAATTCT GTGCTGTATT TGCCATTAAA AAACTGACCT CCTTTAGTTA GTTTTTTGGC 1080

CTAACTTTTG AGGGTCAGTT CAAAATTTGC GACTTTTAAA TGAATTCCAA TATTCAATTA 1140

TTAAGAGTTA ACATGGTGCT TGCCAATAGG AATCATTAGA GGCGAATTGG AAATAGGGTC 1200

ACGTATAATT TTTGCTTCAA GATTAAAGAT ATCTTTAACT AGTTTATCAT TTAGTATATC 1260

TTCAGGCTTT CCCTCTGCAA CAAGTTTACC TTCTTTAATT GCAAATAGGT AATCAGCGTA 1320

TCTTGCTGTT AGATTTATAT CGTGCAAAAT CATGCAAATG GTTGTCTTAT ATTTTTGGTT 1380

TAGATCAGTC AAGAGGTCTA ATAGTTCTAT TTGATATGAG ATATCCAAGT AAGTAGTTGG 1440

CTCATCTAAA AGTAGGATAC TTGTATCTTG GGCTAGGGCT AGAGCTATCC ATACTCTTTG 1500

CCTTTGACCC CCAGAAAGTT CTTCAACTAG GTTATTTGCT AGATCTTCAA CATTGGCCTT 1560

AACCATTGAT CTGTTTATTA TTTCAAGGTC ATCTTTTCCA AGACTCTTAA AAGGCTTTCT 1620

GTAGGGGAAA CGACCACGGC TTACAAGATC AGCTACTGTT ATTGATTCAG GGATTATTGG 1680

AGATTGAGGT AATATAGCTA TGTGTTTTGC TAAATCTTTT TCTTTATAAG AATTAATTGA 1740

TTTATTATCA AGCAATACTT CTCCCTCTAA TGGCTTTATA AGTCGAGACA AGGTTTTAAT 1800

GAGTGTTGAT TTCCCACAAC CATTTGACCC AATAATAACT GATATTTTTT CTTCAGGTAT 1860

TTTTATATTT ATATTTTCCA AGATTATTTT TTCATCATAA CCGCAGGTAA GATTATTTGA 1920

CCACAGACCT TTCATTATAT ATTCCTCCTG TTCATTTTTA TTAGTAAGTA TATTAAGTAT 1980

GGTGAACCTA ACAAGCCAGT TACAACACCT ACTGGATATC TAGCTGGTAA AATATTTTGA 2040

GAGAATATGT CTGATAACAA AACTAGTAAA ATTCCAACCA ATCCAGCTAA TATTGGGCTT 2100

CTTTTCTTGC CAATATTTAA GGCTATGGGA CCAGCTAAAA AAGATATACA AGCTATTGGT 2160

CCTGTAATTG AAGTAGAAAA AGCAGTTAAA GATACAGCGC AAAAAATTAA AACAAGCCTT 2220

GAAAGCTCGG GATTTGCTCC AAGTCCGATT GCTATTTCTT CACCAAGTTC AATAATTTCT 2280

AGTCTTTTAT TAAAAAATAA AACTAATATA GTAGCAATAA TACTTACTAT TAGAACAAGA 2340

GGTATGTCAT CTAACTTTGT AAAAGATAAA GAGCCACTGA GCCATCTCAT AACTTCTTGT 2400

AATTCATATC TTGCTACTTT CAACAATAAA AATGAGGTGC CTGCTCTTGT GACAGCTTGA 2460

AAACCAATAC CTAATATTAT CAGTCTTGCT GCTGAAAAAC CATCTTTTTT AGCTAGTAAA 2520

AATAATATTA AAGATGATGT TAGTCCACAA GTTATTGAAA TAATTCCAGT AGTTAAACTA 2580

TTTGTTTTTA ATACCAATAT GCAAAAGACC GCTGCAATAG ATGAAGAACT TGTGACACCG 2640

ATTATATCAG GACTTGCAAG AGGATTTCTT AACATAGTTT GAAAGATAAA TCCTGCCAAT 2700

CCAAAAGACC AGCCAGCTAT AATTCCTGCT AATAATTTTG GTAATCTAAT TTCCATAATC 2760

GAAAAACTAG CTCCAGGAAC AGTTTCACTA TTTAAGACTT TAATCAAAGT TGAAAAAGAA 2820

TAACTTTCAT CTCCGATAAG TAAAATGAAA AATGATAGAC TGATTATTAT TAATAAAAAT 2880

AGTGAGGAAA ATAGTGTTAT TCTATTTTTT CTTTTTTGAA TACCTATAAT TAAATTTTGC 2940

ATTAGTTATT AACCCCTCTA TTTTTCATAG TTACATAAAT AAGTACTGGA CCCCCGATTA 3000

TTGCAGTAAT TATCCCTACT TCAATTTCAC CTGGTTTACC TAACATACGG CCGATTATAT 3060

CACATATAAG CAAGAGCTCT GCACCTATAA AAGATGAAGA AATGGTCATT GTGCGTATAT 3120

CTTTGCTTAT AAATAAGCCA CAAAAGTGAG GAACTATAAG ACCTACGAAG CCAATAGGTC 3180

CACCAATTGC AGTAATACTT GAACATAAAA GCACACTTGC AATTATTGCA AGTGATCTTA 3240

TCCTATTAAC ATTAACTCCA AGACCAACAG CCATTTCATC ACCCATAGcT AAAGCGTTTA 3300

AATCTGATGA AATAAATATA GCTATCAAGT GACCTAAAAT TATAAAAGGT AGTAGTGTAG 3360

ATATAGAAGA TAATGTAGCT GCTCCAAGGC TACCTATTTG CCAAAATCTA AATTTGTCTA 3420

AGACGTTATT ATTCGGTAAA ATTAAAAAAC TTACAAAACT GCTTAAAGCC ATACTAACAC 3480

1200

AAGTTCCTGA TAAGGCAAGT TTTATAGGGG TAAGGCCTGC TTTTCCGTTA CAGCAATCGC 3540

GTATACAAAA ATTGCACTTA CTAAGCCACC AATGATTGCG 3580

(2) INFORMATION FOR SEQ ID NO: 210:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11378 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

CCAAATTGCT CCACAATTAT TATGGAGTCG TCGTTTGGCA GATGGGCGTG ATATGTGTGC 60

TCAAGAATGG TTGACAGGCA AGATATTGAC CCCCTATGAT ATGAATCGTA AGCAAATCGT 120

CAATATTTTA ACCCGTCTTC ATCGCTCACG TCCGTTGATG ACACAATTGA GTCGTTTGGG 180

CTATGCCATG GAAACACCTG TAGATTTACT ACAGTCTTGG CAGGAAACGG CTCCAGATGC 240

TTTGCGTAAA AATCATTTTA TCAGTGAAGT GATGGCTGAT TTACGTCAGA CTATTCCAGG 300

ATTTAGAGAG GACCATGCGA CCATTGTCCA TGGAGATGTA CGACATAGTA ATTGGATTGA 360

GACAGATAGT GGCTTGATTT ATTTAGTAGA TTGGGATTCG GTTCGCTTGA CCGATCGCAT 420

GTTTGATGTG GCCCATATGC TCTGCCATTA TATTTCAGAA CATCAGTGGA AGGAATGGTT 480

GACCTACTAC GGTTACAAGT ACAATCAAAC GGTATTAAGT AAATTGTATT GGTATGGTCA 540

ATTGTCTTAT TTGAGTCAGA TTTCCAAGTA TTATATGAAC CAAGATTTAG AAAATGTCAA 600

TCGGGAGATT CATGGTTTGC GTCATTTCCG AGACAAGTAT GGAAAGAGAA GATGAGAGTT 660

AGAAATCGTA AAGGGGCAAC AGAATTACTA GAGGCAAATC CCCAGTATGT GGTCCTCAAT 720

CCCTTGGAAG CCAAGGCAAA ATGGCGGGAC TTGTTTGGCA ATGATAATCC CATTCATGTG 780

GAAGTTGGAA GTGGAAAGGG TGCCTTTGTT TCAGGTATGG CCAAGCAAAA CCCTGACATC 840

AACTATATCG GGATTGATAT TCAAAAGTCT GTTTTGAGCT ACGCTTTGGA CAAGGTGCTT 900

GAAGTTGGAG TGCCTAACAT CAAGCTCTTG TGGGTAGATG GTTCTGACTT AACTGACTAC 960

TTTGAAGACG GTGAGATTGA TCGCTTGTAT CTGAACTTTT CAGATCCATG GCCGAAAAAA 1020

CGCCATGAAA AGCGTCGTTT GACCTACAAG ACCTTCTTGG ATACCTTCAA ACGTATCTTG 1080

CCTGAAAATG GAGAAATTCA TTTCAAGACG GATAACCGTG GCTTGTTTGA GTACAGTTTA 1140

GTGAGCTTTT CTCAATATGG CATGAAACTC AATGGTGTCT GGTTAGATTT GCATGCCAGT 1200

GATTTTGAAG GCAATGTCAT GACAGAATAC GAGCAAAAAT TCTCAAACAA GGGGCAAGTT 1260

ATCTACCGAG TTGAGGCAGA ATTTTAAGAG ATAACCTAAA ATTAGGCTGT ACAAGTGCTT 1320

1201

TTGCTTTACA TAAGTTGGCA AACGTGCTAT ACTGATAGTA AGAATATGAA AAGTGAGGCG 1380

GGGAAATATC TTCGCCTCTT GCTTATGAGG AGGTGGACGC AATCGCAACA ATCGTAGAAT 1440

TAGTCAGAGA AGTTGTAGAA CCTGTCATAG AAGCTCCTTT TGAACTCGTG GATATCGAGT 1500

ATGGAAAGAT TGGCAGTGAC ATGATTCTCA GTATTTTTGT AGATAAACCC GAAGAATTAC 1560

CTTGAACGAC ACGGCAGACT TGACAGAAAT TATCAGTCCT GTCCTAGACA CCATCAAGCC 1620

AGATCCCTTC CCAGAACAAT ATTTCCTAGA AATTACCAGT CCAGGTTTGG AACGTCCTTT 1680

GAAAACCAAG GATGCCGTCG CTGGAGCGGT TGGAAAATAC ATCCATGTCG GGCTCTACCA 1740

AGCCATCGAT AAGCAAAAGG TCTTTGAAGG AACCTTGTTG GCCTTCGAAG AGGACGAGTT 1800

GACTATGGAA TATATGGACA AGACGCGTAA GAAAACCGTC CAAATTCCAT ACAGTTTAGT 1860

ATCAAAAGCA CGTTTAGCAG TTAAATTATA GAAAAAGAAA GGATAGCTTT TGAGGATTCA 1920

AAAGTGAAGA AAACATGAGT AAAGAAATGC TAGAGGCCTT CCGCATTTTG GAAGAAGACA 1980

AGGGAATCAA AAAAGAAGAT ATCATCGACG CAGTAGTAGA GTCGCTTCGT TCCGCTTATC 2040

GCAGACGCTA TGGTCAGTCA GACAGCGTAG CTATTGACTT CAACGAAAAA ACAGGTGACT 2100

TTACAGTTTA TACTGTCCGT GAAGTTGTTG ATGAAGTATT TGATAGCCGT TTGGAAATCA 2160

GCTTGAAAGA TGCTCTTGCC ATTAATTCAG CTTATGAACT TGGAGACAAA ATCAAGTTTG 2220

AAGAAGCACC AGCTGAGTTT GGTCGTGTAG CAGCCCAATC TGCCAAACAA ACCATCATGG 2280

AAAAAATGCG CAAgCAAACA CGTGCCATCA CTTACAATAC TTACAAAGAA CATGAGCAAG 2340

AAATCATGTC TGGTACAGTA GAACGCTTTG ACAACCGCTT TATCTATGTC AACCTTGGTA 2400

GCATCGAAGC CCAATTGTCA AAACAAGACC AAATTCCTGG AGAAGTTTTT GCTTCTCATG 2460

ATCGTATCGA AGTTTATGTT TACAAGGTTG AAGACAACCC TCGTGGTGTG AACGTCTTTG 2520

TTAGCCGTAG TCATCCAGAA ATGATCAAAC GTTTAATGGA GCAAGAAATT CCAGAAGTTT 2580

ATGATGGAAC TGTTGAAATC ATGAGCGTGG CTCGTGAAGC AGGTGACCGT ACGAAGGTTG 2640

CTGTTCGTAG CCACAATCCA AACGTGGATG CTATCGGTAC AATCGTTGGA CGTGGTGGTG 2700

CTAATATCAA GAAGATTACT AGCAAATTCC ACCCAGCTCG TTACGATGCT AAAAATGACC 2760

GCATGGTACC AATCGAAGAA AATATCGATG TTATCGAGTG GGTAGCAGAT CCAGCTGAAT 2820

TTATCTACAA TGCCATCGCT CCTGCTGAGG TTGACCAAGT TATCTTTGAT GAAAACGACA 2880

GCAAACGTGC CTTGGTGGTT GTTCCAGATA ACAAGCTTTC TCTTGCCATT GGTCGTCGTG 2940

GACAAAACGT GCGCTTGGCG GCTCACTTGA CTGGTTACCG TATCGATATC AAGTCTGCTA 3000

GCGAATTTGA AGCCATGGAA GACGCTGCTT CAGTAGAGTT GGAAGTAGAA AACGATACTG 3060

TAGAAGAATA AAAGCTGCTA GAGGAGGGAA AGATGAAAAC AAGAAAAATC CCTTTGCGCA 3120

AGTCTGTTGT GTCTAACGAA GTGATTGATA AGCGTGATTT GCTCCGCATT GTCAAGAACA 3180

AGGAAGGACA AGTCTTTATT GATcCTACGG GCAAGGCCAA TGGCCGCGGC GCTTATATCA 3240

AACTAGACAA TGCAGAAGCC CTAGAGGCGA AAAAGAAGAA GGTCTTTAAC CGCAGCTTTA 3300

GCATGGAAGT GGAAGAAAGC TTTTATGACG AGTTGATCGC TTATGTGGAT CACAAAGTGA 3360

AAAGAAGAGA GTTGGGACTT GAATAAGCAA AAGATAAGTA ATCTCTTGGG GCTTGCTCAG 3420

CGAGCAGGGC GCATCATATC GGGTGAAGAA TTGGTGGTCA AGGCCATTCA AGACGGCAAG 3480

GCCAAGTTGG TCTTTCTAGC TCATGATGCT GGACCCAATC TGACCAAGAA GATTCAAGAT 3540

AAAAGTCATT ATTATCAAGT AGAAATTGTA ACCGTGTTTT CAACACTGGA ATTAAGCATA 3600

GCAGTCGGGA AATCGAGAAA GGTTTTGGCT GTAACAGATG CTGGATTTAC AAAGAAAATG 3660

AGGTCTCTTA TGGAATAGAA GAGGAGGACA TGATTTGTCT AAGAAAAGAT TGTACGAAAT 3720

CGCAAAAGAA CTTGGAAAAG AAAGTAAAGA AGTTGTAGCG CGTGCAAAAG AGTTGGGCTT 3780

GGATGTGAAA AGCCACTCAT CAAGTGTGGA AGAAGCTGTC GCTGCAAAAA TTGCTGCCAG 3840

CTTTAAGCCT GCAGCTGCTC CGAAAGTAGA AGCAAAACCT GCAGCCCCAA AAGTAAGTGC 3900

AGAAAAGAAA GCCGAAAAAT CTGAGCCAGC TAAACCAGCT GTAGCTAAGG AAGAGGCAAA 3960

ACCTGCAGCC CCAAAAGCAA GTGCAGAAAA GAAAGCCGAA AAGTCTGAAC CAGTAAAACC 4020

AGCTGTAGCC AAGGAAGAGG CAAAACCAGC TGAGCCAGTC ACTCCGAAAA CAGAAAAAGT 4080

AGCGGCTAAA CCGCAAAGTC GTAATTTCAA GGCTGAGCGT GAAGCACGTG CTAAAGAGCA 4140

GGCAGAGCGA CGCAAGCAAA ATAAGGGCAA TAACCGTGAC CAACAACAAA ACGGAAACCG 4200

TCAGAAAAAC GACGGCCGTA ATGGTGGAAA ACAAGGTCAA AGCAACCGCG ACAATCGTCG 4260

CTTTAATGAC CAAGCTAAGA AGCAGCAAGG TCAGCAAAAA CGTAGAAATG AGCGCCGTCA 4320

GCAAGAGGAT AAACGTTCAA ATCAAGCGGC TCCACGTATT GACTTTAAAG CCCGTGCAGC 4380

AGCCCTAAAA GCAGAGCAAA ATGCAGAGTA CGCTCGTTCA AGTGAGGAAC GCTTCAAGCA 4440

GTATCAGGCT GCTAAAGAAG CCTTGGCTCA AGCTAACAAA CGCAAGGAAC CAGAGGAAAT 4500

CTTTGAAGAA GCGGCTAAGT TAGCTGAACA AGCACAGCAA GTTCAAGCAG TGGTTGAAGT 4560

CGTCCCTGAG AAAAAAGAAC CTGCAGTGGA TACACGTCGT AAAAAACAAG CTCGACCAGA 4620

CAAAAATCGT GACGATTATG ATCATGAAGA AGATGGTCCT AGAAAACAAC AAAAGAATCG 4680

AAGTAGTCAA AATCAAGTGA GAAATCAAAA GAATAGTAAC TGGAATAACA ACAAAAAGAA 4740

CAAAAAAGGC AATAACAAGA ACAACCGTAA TCAGACTCCA AAACCTGTTA CGGAGCGTAA 4800

ATTCCATGAA TTGCCAACAG AATTTGAATA TACAGATGGT ATGACCGTTG CGGAAATCGC 4860

AAAACGTATC AAACGTGAAC CAGCTGAAAT TGTTAAGAAA CTTTTCATGA TGGGTGTCAT 4920

GGCCACACAA AACCAATCCT TGGATGGGGA AACAATTGAA CTCCTCATGG TGGATTACGG 4980

TATCGAAGCC AAACAAAAGG TTGAAGTGGA TAATGCTGAC ATCGAACGTT TCTTTGTCGA 5040

AGATGGTTAT CTCAATGAAG ATGAATTGGT TGAGCGTCCA CCAGTTGTTA CTATCATGGG 5100

ACACGTTGAC CACGGTAAAA CAACCCTTTT GGATACTCTT CGTAACTCAC GTGTTGCGAC 5160

AGGTGAAGCA GGTGGTATTA CTCAGCATAT CGGTGCCTAC CAAATCGTGG AAAATGGTAA 5220

GAAGATTACC TTCCTTGATA CACCAGGACA CGCGGCCTTT ACATCAATGC GTGCGCGTGG 5280

TGCTTCTGTT ACCGATATTA CGATCTTGGT CGTAGCGGCA GATGACGGGG TTATGCCTCA 5340

GACTATTGAA GCCATCAACC ACTCAAAAGC AGCTAACGTT CCAATCATCG TAGCTATTAA 5400

CAAGATTGAT AAACCAGGTG CTAACCCAGA ACGCGTTATC GGTGAATTGG CAGAGCATGG 5460

TGTGATGTCA ACTGCTTGGG GTGGAGATTC TGAATTTGTT GAAATTTCGG CTAAATTCAA 5520

CCAAAATATC GAAGAATTGT TGGAAACAGT CCTTCTTGTG GCTGAAATCC AAGAACTCAA 5580

AGCAGACCCA ACAGTTCGTG CGATCGGTAC GGTTATCGAA GCGCGCTTGG ATAAAGGAAA 5640

AGGTGCGGTC GCAACCCTTC TTGTACAACA AGGTACCTTG AATGTTCAAG ACCCAATCGT 5700

TGTCGGAAAT ACcTTCGGTC GTGTCCGTGC TATGACCAAC GACCTTGGTC GTCGTGTTAA 5760

AGTTGCTGGA CCATCAACAC CAGTCTCTAT CACAGGTTTG AACGAAGCAC CGATGGCGGG 5820

TGACCACTTT GCCGTTTACG AGGATGAAAA ATCTGCGCGT GCAGCAGGTG AAGAGCGTGC 5880

CAAACGTGCC CTCATGAAAC AACGTCAAGC TACCCAACGT GTTAGCCTTG AAAACCTCTT 5940

TGATACCCTT AAAGCTGGGG AACTCAAATC TGTTAATGTT ATCATCAAGG CTGATGTACA 6000

AGGTTCTGTT GAAGCCCTTT CTGCCTCACT TCAAAAGATT GACGTGGAAG GTGTCAAAGT 6060

GACTATCGTC CACTCAGCGG TCGGTGCTAT CAACGAATCA GACGTGACCC TTGCCGAAGC 6120

TTCAAATGCC TTTATCGTTG GTTTCAACGT ACGCCCTACA CCACAAGCTC GTCAACAAGC 6180

AGAAGCTGAC GATGTGGAAA TCCGTCTTCA CAGCATTATC TACAAGGTTA TCGAAGAGAT 6240

GGAAGAAGCT ATGAAAGGGA TGCTTGATCC AGAATTTGAA GAAAAAGTTA TTGGTGAAGC 6300

GGTTATCCGT GAAACCTTCA AGGTGTCTAA AGTGGGAACT ATCGGTGGAT TTATGGTTAT 6360

CAACGGTAAG GTTGCCCGTG ACTCTAAAGT CCGTGTTATC CGTGATGGTG TCGTTATCTA 6420

TGATGGTGAA CTCGCAAGCT TGAAACACTA TAAAGACGAC GTGAAAGAAG TGACAAACGG 6480

TCGTGAAGGT GGATTGATGA TCGACGGCTA CAATGATATT AAGATGGATG ATGTGATTGA 6540

GGCGTATGTC ATGGAAGAAA TCAAGAGATA AGATTTTTTG CTCCTTTCTT AGGTGGTGAG 6600

GGACGCAAGC AAACCGATGG TTTCATTGCT TATTTTTGAG CCTAGGGTCT CAAAAATCCC 6660

CTGTGATGGG ACTGATAAAT CAGTTCCATC ACTTTCACCA CGGCGAAAGA AGCAGATGAC 6720

TTCAAATTGA ACTTCGTTTC AATTTAAACT GAAAATCAAG AAGTTTAAAA TAGCTAGGTC 6780

TGCTGGCCTA GCTTTTGGTT CAAAGTAGAG AAAGGAATAT CATGGCAAAT CATTTCCGTA 6840

CAGATCGTGT GGGCATGGAA ATCAAGCGTG AAGTCAATGA GATTTTGCAA AAGAAAGTCC 6900

GTGATCCACG TGTCCAAGGT GTGACCATCA TAGATGTTCA GATGCTGGGT GACTTGTCTG 6960

TTGCCAAGGT TTATTACACC ATTTTGAGTA ACCTTGCTTC GGATAACCAA AAAGCCCAAA 7020

TCGGGCTTGA AAAAGCAACT GGTACCATCA AACGTGAACT TGGTCGCAAT TTGAAATTGT 7080

ACAAAATCCC AGATTTGACC TTCGTCAAAG ACGAGTCCAT CGAGTATGGA AACAAGATTG 7140

ACGAGATGCT ACGCAATCTG GATAAGAACT AAAGAAGAGG GGTTGCCCCT CTTTTTTGGT 7200

GGAGGAAAAT AGGTTGAATT TGAAATGGAA AAATATTCTT TTATAATAGA TTGAAACTAG 7260

AATAGTACGC CTCTACTTCT AAAATATTGT TAGAAATCGA TTTGACTGTC CTGATCGATT 7320

TGTCCTGTTC TTGTTTCATT TTAATATAAA AAAGGGATTC TGTATTTTTT AATGTTATCT 7380

AATTAGAAAA TGCTTTTTTT GTAGGAAATA TAATATGATA AGGTGCAAAA AAGAAATAAG 7440

GAGTTTGTAT ATGGCTGAAC AAGACTTAGC TATGCAAGTA TTGCAACAAG TGGTGAAACT 7500

ACCTGTTGTT AAGGTTGATC GTTCGAAATT TTTAGTGGAT AAGTTTTCCA AAGAATTGGA 7560

TCCAAAAGAT ATTCCTACCT TATTGGAACA AGGTCCAACG ACTCTTCTAT CTCAAGAAAT 7620

ATTAGATCGT GTAGCTAATG CTTGTATTCG GGACAATGTA TTATTAGCGA GTGGGACTTC 7680

TGTTTTGGCA GGATTACCTG GAGGGCTTGC TATGGCAATT ACCATTCCAG CTGATGTGGC 7740

TCAATTTTAT GCTTTCTCTC TGAAATTGGC TCAAGAATTA GGTTATATTT ATGGTTATGA 7800

GGATCTTTGG GCTTCACGAG AGGAGTTGAG TGAAGATGCT CAAAATACCC TCTTGCTTTA 7860

TCTAGGCGTA ATGTTAGGGG TGAATGGAAC CGCTGCTTTG CTACGTGTTG GTAGTATAAC 7920

AATTGCCAAA CAGGTAATGA AAATAGTGCC TAATAAAGCT TTAACAAAGA CGCTTTGGTA 7980

CCCTATTTTG AAAAAAGTCT TAAAAATATT TGGTGTGAAT CTTACCAAGG GAGGGTTGGC 8040

CAAAGGAATG GGGAAATTTA TTCCTATCTT GGGTGGTATC ATTTCAGGTG GTTTAACCTT 8100

TGCAACTATG AAACCAATGG GGGAAAGCTT GCAGAAAGAA TTATCCAAGC TAGTCAACTA 8160

TAGTGAAGTT CAATATCAAG AAGATGTTGA AACAATCCGA AAAGAGGCTG AAATCATCAA 8220

AGGAGAGTAA TATGAATCCT ATCAAAGCTT TTGCTAAAAT TTATGGTAAT TACTTTTTGA 8280

CCGTGCAAGG TGTAAAAGTG ATGAAAACGA TAAAGAAAGC TGACCATGTC GTTGTTGGTC 8340

TGGGGAAACT TTTTATTGCC GACAAGTTAA TGGATACGGC TCGGTGGCTC ATTAAGCCAG 8400

AGGAGAGAGA ATGAAATTTT TTTGGTCTTC TTGCTATTCT TTTTATCAAA CCGATTATTG 8460

GGATTGTGAA ATTCTTTTGG ATGATCATCT CTTTTGCAGT CCAATTGCTG TTTTACAAGA 8520

TAGTGTTTAA GATATTGGAT TGGCTCTTTA AACTTATCTA GATGGTAATC CAAGTTGCAG 8580

AGAACTAGCA GGAACTCCAC TGCTAGTTTT TTATTCTCTT TCCATATGGT ATAATATAAG 8640

CAGTAAAATC ATTTTATACT CTTCGAAAAT CTCTTCAAAC CACGTCAGCT TCACCTTGCA 8700

GTATATATGT TACTGACTTC GTCAGTTCTA TCCACAACCT CAAAACGGTG TTTTGAGCTG 8760

ACTTCGTCAG TTCTATCTAC AACCTCAAAA CACTGTTTTG AGCAACCTGC GGCTAGCTTC 8820

CTAGTTTGCT CTTTGATTTT CATTGAGTAT TAGAACATAC AATGGAGGTC GTCATGGACA 8880

ATATCATCGA TGTGTCAATT CCTGTTGCAG AAGTGGTGGA CAAGCATCCA GAAGTCTTGG 8940

AAATTCTAGT GGAGTTGGGT TTTAAACCCC TTGCCAATCC CTTAATGCGC AATACAGTTG 9000

GTCGTAAAGT ATCACTTAAA CAGGGTTCTA AGCTAGCAGG AACTCCTATG GACAAGATTG 9060

TACGCACACT GGAAGCGAAT GGCTACGAAG TGATTGGATT AGACTAATGA CAGATGAACG 9120

GATTCATATC CTACGGGATA TTTTGTTAGA ATTGCACAAT GGCGCCTCTC CTGAGTCGGT 9180

TCAAGATCGC TTTGATGCGA CCTTTACGGG CGTGTCAGCC ATCGAGATTT CCCTTATGGA 9240

GCACGAGCTG ATGAACTCGG ATTCGGGCGT CACTTTTGAA GATGTTATGG AACTCTGTGA 9300

TGTCCATGCC AATCTTTTTA AAAATGCTAT CAAAGGTGTC GAAGTTTCAG ATACTGAGCA 9360

TCCAGGTCAC CCAGTTCGTG TCTTCAAAGA AGAAAATCTG GCTCTCCGTG CGGCCTTGAT 9420

TCGCATTCGT AGATTGTTAG ATACCTATGA GTCTATGGAA GACGAGGAAA TGCTGGCGGA 9480

GATGCGTAAG GGTTTGGTGC GTCAGATGGG ACTTGTGGGT CAATTTGACA TCCATTACCA 9540

ACGTAAGGAA GAACTCTTCT TTCCTATCAT GGAGCGCTAT GGACACGATT CACCTCCCAA 9600

AGTTATGTGG GGAGTGGATG ATCAGATTAG GGAACTCTTT CAAACAGCTC TAACGACAGC 9660

CAAGTCACTA CCAGAAGTGT CAATTAGCAG TGTAAAGGAA GCTTTTGAAG CTTTTGCGAC 9720

AGAGTTTGAA AGTATGATTT TCAAGGAAGA GTCCATCCTC CTCATGATTC TCCTTGAGTC 9780

TTTTACTCAG GATGACTGGC TTCAGATTGC GGAGGAGAGC GATGCCTATG GCTATGCCAT 9840

CATCCGTCCG TCAGAGAAAT GGGTGCCAGA ACGACAGAGC TTTATTGAGG AAAAGATTGC 9900

AGAGGAGCCT GTACAGCTAG ATACGGCAGA AGGTCAAGTT CAACAAGTCA TAGATACGCC 9960

AGAAGGCCAT TTTACCATTA CCTTTACCCC TAAGGAAAAG GAAGCTGTGC TGGACCGCCA 10020

TAGTCAACAG GCTTTTGGTA ATGGCTATCT TTCAGTCGAG CAGGCCAATC TCATCCTCAA 10080

TCATCTCCCT ATGGAGATTA CCTTTGTCAA TAAAGAAGAT ATTTTCCAGT ATTACAATGA 10140

CAATACGCCA GCTGATGAGA TGATTTTCAA ACGGACGCCG TCCCAAGTCG GGCGCAATGT 10200

CGAACTCTGC CATCCGCCTA AGTACTTGGA CAAGGTCAAA ACTATCATGA AGGGGCTTCG 10260

TGAGGGAAGC AAAGACAAGT ATGAAATGTG GTTCAAGTCT GAGTCGCGAG GTAAGTTTGT 10320

CCACATCACC TATGCTGCAG TACACGATGA AGACGGAGAA TTCCAAGGAG TGTTGGAGTA 10380

TGTTCAGGAT ATCCAGCCCT ACCGTGAGAT TGATACGGAC TATTTTCGTG GATTAGAATA 10440

AGGAGAAAAA ATGAGTTACG AACAAGAATT TATGAAGGAA TTTGAAGCTT GGGTCAATAC 10500

CCAAATCATG ATTAACGACA TGGCGCACAA GGAAAGCCAA AAAGTTTACG AAGAAGACCA 10560

GGACGAGCGT GCCAAAGATG CCATGATTCG CTACGAGAGT CGCTTGGATG CTTATCAGTT 10620

CTTGCTTGGT AAGTTTGAAA ACTTCAAAGT AGGCAAGGGA TTCCATGATT TGCCAGAAGG 10680

CTTGTTTGGT GAGCGAAATT ATTAAACGAG AAAGATTCTT GATTTTTCAC TAAAATCTTG 10740

ATAGAATGTT TATGTTAAAT CCTTGTCAGA GCAGGGATTT TTTATTGAAA GGATTTTATC 10800

ATGTCAAAGA AACTCAATCG TAAAAAACAA TTACGAAATG GCCTCCGTCG CGCAGGTGCC 10860

TTTTCAAGTA CGGTGACTAA GGTTGTAGAT GAGACAAAAA AAGTCGTGAA GCGTGCAGAA 10920

CAGTCAGCAA GCGCAGCTGG TAAGGCTGTT TCTAAAAAAG TTGAACAAGC AGTAGAAGCT 10980

ACCAAAGAGC AAGCTCAAAA AGTAGCTAAT TCTGTAGAAG ATTTTGCAGC AAATTTGGGT 11040

GGACTTCCAC TTGATCGTGC CAAGACTTTC TATGATGAAG GAATCAAGTC TGCTTCAGAT 11100

TTCAAAAACT GGACTGAAAA AGAACTCCTT GCCTTGAAAG GAATCGGCCC AGCTACCATC 11160

AAGAAATTGA AAGAAAATGG CATCAAGTTC AAGTAATTTT TCTTGAGCCT TGCATTTCCG 11220

AAAAAATCTT GCTACAATAG AGCCATTAGA GGTGTTTTGA ATCCCACATT TTACAGAAAG 11280

TGGCGGCGCT GAGAAGTCCA CAAATGTGTC AAAACTGGTT GCTAATGGAT GAAAAATTGA 11340

AATAAAAGTG TCTTTTTGCT TTAAAGACGA GAGTTGCG 11378

(2) INFORMATION FOR SEQ ID NO: 211:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4156 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

CCGCGAGCCA CGGCGAATTT GCTGCGGGTA TTCATCAGTC AGGATCTATG ATCTTTGGTG 60

AACAAGAAAA GGTTCAAGTT GTGACCTTTA TGCCAAATGA AGGTCCTGAT GATCTATACG 120

CTAAGTTTAA TAACGCTGTT GCTGCATTTG ACGCAGAAGA TGAGGTTCTA GTTTTGGCTG 180

ACCTTTGGAG TGGTTCTCCA TTTAACCAAG CTAGTCGCGT GATGGGAGAA AATCCTGAGC 240

GTAAGTTTGC CATCATCACA GGACTTAACT TACCGATGTT GATTCAAGCC TACACAGAGC 300

GCCTCATGGA CGCTGCTGCA GGTGTAGAAA AAGTCGCTGC TAATATCATT AAAGAAGCCA 360

AAGATGGCAT CAAAGCTCTT CCAGAAGAGC TAAATCCAGT CGAAGAAGTT GCAAGCGCTG 420

CAGCTGCTCC AGTTGCCCAA ACTGCTATCC CAGAAGGAAC TGTTATCGGA GACGGTAAAT 480

TGAAAATCAA TCTTGCCCGT CTTGACACAC GTCTACTTCA CGGTCAGGTT GCAACTGCTT 540

GGACTCCAGA TTCAAAAGCA AATCGTATCA TCGTTGCTTC AGATAACGTG GCTAAAGACG 600

ACCTTCGTAA AGAATTGATT AAACAAGCAG CTCCAGGTAA TGTCAAGGCT AACGTGGTTC 660

CAATTCAAAA ACTGATTGAG ATTTCAAAAG ACCCACGTTT TGGAGAAACA CATGCCCTTA 720

TCTTGTTTGA AACACCTCAA GATGCCCTTC GTGCCATCGA AGGCGGCGTG CCAATCAAGA 780

CTCTTAATGT TGGTTCTATG GCTCACTCAA CAGGTAAAAC ATTGGTCAAT ACCGTTTTGT 840

CTATGGACAA AGAAGACGTT GCTACATTTG AAAAAATGCG TGACTTGGGT GTTGAATTTG 900

ATGTCCGTAA AGTACCAAAT GATTCTAAAA AAGATTTGTT TGACTTGATT AACAAAGCCA 960

ATGTCAAATA AGCCATTATT TATGAAAGGA TTTTAAACAT GTCTATTATT TCTATGGTTT 1020

TAGTAGTCGT TGTAGCCTTC TTTGCAGGTC TTGAAGGCAT CCTCGACCAG TTCCAATTTC 1080

ACCAACCACT TGTAGCCTGT ACCCTTATTG GGCTTGTAAC AGGTCACTTG GAAGCAGGGA 1140

TTATCCTCGG TGGATCGCTT CAAATGATTG CCCTTGGTTG GTCAAATATC GGTGCTGCTA 1200

TCGCTCCTGA TGCTGCACTT GCTTCTGTCG CTGCTGCCAT TATCATGGTT CTTGGTGGTG 1260

ACTTTACCAA GACTGGTATC GGTGTTGCCC AAGCGGTTGC TATCCCTCTT GCTGTAGCTG 1320

GACTTTTCTT GACAATGATT GTTCGTACAA TTTCAGTTGG TTTGGTTCAT ACTGCAGATG 1380

CTGCCGCTAA AAAAGGTGAC TTCGGCGCTG TGGAGCGTGC GCATTTCATC GCGCTACTTT 1440

TCCAAGGACT TCGTATCGCG CTTCCTGCAG CTCTTCTCCT TATGGTACCA ACTGAAACTG 1500

TACAAAGTAT CCTTAGTGCC ATGCCAGACT GGCTCAAAGA TGGTATGGCT ATCGGTGGTG 1560

GTATGGTCGT TGCCGTTGGT TACGCCATGG TTATCAACAT GATGGCAACT CGTGAAGTAT 1620

GGCCATTCTT CGCTCTTGGT TTCGTTCTCG CTGCTGTGTC AGATATTACT CTAATCGGAT 1680

TCGGTGCTAT CGGCGTTGCT ATCGCTCTTA TCTACCTTCA CCTTTCTAAA ACTGGTGGAA 1740

ATGGTGGCGG AGGAGCCGCA ACTTCTAACG ACCCAATCGG CGATATCCTA GAAGACTACT 1800

AAGATAAGAA AGGACTGAAA ACATCATGAC TGAAAAACTT CAATTAACTA AATCAGATCG 1860

TAAAAAAGTT TGGTGGCGTT CAACCTTCTT ACAAGGGTCT TGGAACTTTG AACGGATGCA 1920

AAACTTGGGC TGGGCTTATA CACTCATTCC AGCTATCAAA AAACTCTATA CTAAAAAAGA 1980

AGATCAAATC GCTGCTCTTG AGCGTCACCT TGAGTTCTTC AACACTCATC CATACGTAGC 2040

TGCTCCAGTC ATGGGGGTTA CTCTTGCGCT TGAAGAAGAA CGTGCTAACG GTGTGGAAAT 2100

CGATGACGCT GCTATCCAAG GGGTTAAAAT CGGTATGATG GGACCTCTTG CTGGTATCGG 2160

TGACCCAGTA TTCTGGTTTA CAGTACGCCC AATCCTTGGA TCTCTCGGTG CTTCACTTGC 2220

CCTTACTGGC AATATCTTGG GGCCACTCCT CTTCTTTGTT GCATGGAACT TGATTCGTAT 2280

GTCATTCTTG TGGTATGTTC AAGAGATTGG ATACAAGGCT GGATCAGAAA TCACTAAAGA 2340

TATGTCTGGT GGTATCCTTC AAGATATCAC TAAAGGAGCT TCTATCCTTG GGATGTTCAT 2400

TCTTGCTGTC CTTGTTCAAC GCTGGGTAAA TATTAAATTT GCTTTCGATG TTTCTAAAGT 2460

TCAACTAGAT GAAAAGGCTT ATATCCATTG GGATAAATTG CCAGAAGGGT CTAAAGGTAT 2520

CCAAGAAGCA TTCGCACAAG TAGGACAAGG ATTGTCTCAA ACTCCTGAAA AAGTTACTAC 2580

TTTCCAACAA AACTTGGATA TGTTGATTCC TGGATTATCA GGACTACTCC TTACTTTACT 2640

TTGCATGTAC TTACTTAAGA AAAAAGTATC TCCAATCACT ATTATCCTTG CCCTCTTCGC 2700

AGTGGGTATT GTGGCACATG TTCTTCACAT CATGTAATCA AGCAACTAAA AAGGAACCAG 2760

GTTCTAAAAT CTGATTCCTT TTTTCTATGC TTTTATTCAG CCAAGGCTCC CATTGGATCC 2820

CATGGTGCAA GTACGATTGG TTCTGCTCCA TAGGCAGCTT GTTCTTCTGC TGTCAGCAAT 2880

TCCTTACGAA CAACGATTTG GTATGTGTAT TCGTCCATCC AAGCGTCTGA GGCAACAAAG 2940

TAACCATCTG TACCGACCTT GTCTCCCCAT GAGTTTTCAA CCTTCCACTT GGTTGATTTA 3000

CCATTTTCGT CCAAGTCAAC ACCTGTCAAG ACCATGGCGT GGGTCATCAA GCTTTCACTA 3060

TAGTCCAAAC GTCCAGCCTT GTCTTGAGTA AGTTTAATGT CCATGCTTGA TTCAAAGTCA 3120

TAAACATCTG TCGCAAGGAT GCCAGCTTAC GGTTGCTGAG CTGGCCGACA TCAGAACCAA 3180

ACCAAACAGT CTCACCTGCT TGCATTTGGG CAATCGCCAA TTCTTTCAAG CGCTCCATTG 3240

GAACGTTGAT GTAGCGAACT GCACGGCTAC CAACCACATT CCCCAACATC TCAACTGTGT 3300

AAGATTTTCC GTAAGGTTTA TCAGCAGTTG GAGCATTGAT AACAGAAACG TAGTCTTCTA 3360

AAGGAAGATT GACATATTTC TTGTAAAACT CTTGTGGTGT GATTCCTTTT TCACTTTTGT 3420

AGTTGTTATC TTTATCGCGA TAAGCAAAGT CAAACTTGCG TGGTGGAAGT CCTAATGACA 3480

TAGCAAGAAA GTTAAAGATT TCTTGCAAGA GGTCTTCTTT CTTAGCTTGA ACAGTCGCTT 3540

GATCTGCACC AGAAACAAGC AAGTCACGCA AGATTTGAGC ATCTTGACGA AGCAATTTAT 3600

TAAGGATCGC ATTTAGCTCA CGACTGCTGC TAGATGAAAC AGACTCAGGA TAAACTGACT 3660

TAGGCACGAC ACCGTATTTT TCAAAGAGGG AAACGACCAT ATCCCATTGA CCGCCATCTT 3720

GTTGAGGTGT TTGGAGTAAG AAGCTAACtT GCGGCTAGTC AATTCTTGGT CTGAAGTCGC 3780

AATGACTTGC TCCAAGAACC AGTTTGATTT CTCATACTTA TCCCAGAAGA AAGTGTGGGC 3840

TTGTGACAAC TCAAAGTTCT CCAATTTGTA TTGCGAGATG AGTTTGTGGC GGAAGGTGTT 3900

GAGAGCCGCA AACATCCAGC AACGACCAGA CGCTTTCTGG TTAGTGACCT TGTCCTTGGT 3960

TAAATCCAAT GAGAAAACAG GTGTGTTGTC TACATGGCTT TGGCGACGTT CCAGAGCTGC 4020

AAAAATTCCG TTGTGGCTGG CAGCATTTTC AATCGCTTGG TATTTTACAT TTGCTTCATA 4080

GTTGGCAAAT AGTTTATCAG TAAATGATTC TTGAATCGCG TTCATAGATT CCTCCTTTTA 4140

GTCTACAGTG TATTGG 4156

(2) INFORMATION FOR SEQ ID NO: 212:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3902 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 212:

AAAAACAACA AAATAAAACA AAAACAAAAA TATCGAGGTT TATTTTCAAA ACTTTCGATA 60

TTTTTATTAA GTTATTATTT TGTTGTTTCT AGTTTACTTT TTGATGGTTA AGAGTGGTGG 120

AGAATTATAC TCAATGAAAA TCAAAGAGCA AACTAGGAAG CTAGCCGCAG GCTGTACTTG 180

AGTACGGCAA GGCGAAGCTG ACGTGGTTTG AATTTGATTT TCGAAGAGTA TTAGTGCAAA 240

CCGTAGTTGT AGTCATCATC TTGCATGGCT TCAACTTCGC CAAGAAGGTA ACCATTTCCG 300

ACTTGAGAGA AGAAGTCATG GTTGGAAGTT CCTGTTGAAA TACCGTTCAT AACGATTGGG 360

TTGACATCTT CAGCTGAATC TGGGAAAAGT GGATCTTGTC CCATGTTCAT GAGAGCTTTA 420

TTGGCATTGT AGCGAAGGAA GGTTTTAACC TCTTCAGTCC AACCAACACC GTCATAAAGA 480

CTCTCTGTGT AGCCTTCTTC ATTTTCATAA AGAGTATAGA GTAGGTCGTA CATCCATTCT 540

TTGAGTTTTT CTTGCTCTTC TTCAGGTAAT TCATTGAAAC CAAGTTGGAA TTTGTAACCA 600

ATGTAGGTTC CGTGAACAGA CTCGTCACGA ATAATCAATT TAATGATTTC TGCAACGTTG 660

GCAAGTTTGT TGTTACCGAG ATAGTAGAGG GGAGTGAAGA AACCAGAGTA GAAGAGGAAG 720

GTTTCGAGGA AGACGCTGGC AACTTTCTTT TCAAGTGGGC TGCCGTTTAG GTAGATTTCG 780

TTGACAATCT CAGCCTTCTT TTGTAGGTAA GGATTGGTAT TGGTCCATTC GAAAATTTCT 840

TCAATCTCAG CCTTAGTATT CAAGGTAGAA AAGATTGATG AGTAAGATTT AGCGTGGACA 900

1210

GATTCCATAA ATTGGATGTT ATTGAAGACA GCTTCCTCAT GTGGTGTACG GATGTCTGCG 960

CGAAGGGCTT GAACCCCAGT TTCAGATTGC ATAGTGTCAA GAAGGGTTAA ACCACCAAAA 1020

ACTTTTCCGA CCAAGTCTTT CTCTTTGTTA GATAGCTTTC TCCAGTCATC CAAGTCGTTT 1080

GATAAGGGAA TACGTGTATC GAGCCAAAAT TGCTCCGTCA GTTTTTCCCA AGTTGATTTG 1140

TCGATGACAT CTTCGATGGC ATTCCAGTTA ATGGCTTTGT AGTAAGTTTC CATTTAAAAT 1200

CTCTTTCTGT GTTTAGTATT GCGAACTCAC AATTATTTCT ACTTTACCAT AATTCTATAG 1260

GAGTATCGCA CAAAAAGTCG GAAGCCCGAC TTTTAAAATG TTACATAAAT TATGTTATGA 1320

CATAGTAGAT TTGATTTTAT CAGTGCTGCT TAGGGAAAAA TAGTGTTTCT ATGCTAGAAA 1380

CTAAATCACA CAGCTTTCAC ATTGGTTGGC GCCGACTTCT CCACCGTCAT CTGTAAAGGT 1440

ACGGACGTAG TAGATAGACT TGATTCCCTT GTTAAAGGCA TAGTTACGAA GGATGGACAA 1500

GTCACGTGTC GTTTGTTTAT TTTCCCTCTT CCATTCGTAA AGGCCTTTTG GAATGTCACT 1560

GCGCATGAAG AGGGTGAGTG AAAGTCCTTG ATCCACGTGT TCAGTCGCAG CAGCGTAAAC 1620

ATCGATGACT TTACGCATAT CCATATCGTA GGCAGAAGTG TAGTAAGGAA TGGTTTCTGT 1680

AGACAAGCCA GCAGCAGGGT AATAGATTTT ACCAATTTTC TTCTCTTGGC GTTCTTCGAT 1740

ACGTTGCGTA ATCGGGTGGA TAGAAGCAGA AACGTCGTTG ATATAGCTGA TAGAACCATT 1800

TGGCGCTACA GCAAGGCGAT TTTGGTGGTA AAGACCATCT TCTTGAACCT TGTCGCGAAG 1860

TTCAGCCCAA TCAGCAACAC CAGGGATAAA GACATTTTTG AAGAGTTCTT TAACACGGTC 1920

TGATGTTGGA ACAAATTCAC CAGTTACATA CTTGTCAAAG TAACTTCCGT TAGCATAGTC 1980

TGATTTTTCA AAGTTGTGGA AGGTAATACC ACGTTCACGT GCAATATTGT TTGACTCTAC 2040

CAAGGTCCAG TAGTTCATAA GCATAAAGTA GATGCTTGTA AATTCAACAG ACTCAGGTGA 2100

ACCATATTCA ATGAGTTGTT GGGCAAGGTA GCTGTGCAGT CCCATGGCAC CGAGACCAAA 2160

GGTGTGGGCT TGGCTATTTC CATGGTCAAT CGTTGGTACA GCTACGATAT GTGAACTATC 2220

TGTAACGAAA GTAAGGGCAC GAACCATAGC ACGGATAGAA CGACCAAAAT CAGGTGAAGT 2280

CATCATGTTA ACCACGTTGG TTGAACCCAG GTTACATGAA ACATCTGTTC CCATTTGAAG 2340

GAATTCTTGA GCATCGTTGA TCAAGCTTGG TTCTTGAACT TGAAGAATCT CAGAACACAA 2400

GTTACTCATG ATAATCTTTC CATCAACAGG ATTTGCACGG TTAGCCGTAT CGATGTTGAC 2460

TACATAAGGA TAGCCAGACT CTTGTTGCAA TTTAGAGATT TCAGTTTCCA AATCCCGCGC 2520

CTTGATTTTT GTCTTGCGAA TATTTGGATT TGCGACCAAT TCATCGTATT TTTCAGTAAT 2580

GTCGATGTAA TTGAATGGCA CACCGTATTC TTTTTCTACA GAGTAAGGGC TGAAGAGGTA 2640

CATTTCTTCA TTTTTACGAG CCAATTCGTA GAATTTATCA GGTACTACAA CACCAAGTGA 2700

TAGAGTCTTG ACACGTACTT TTTCATCAGC GTTTTCTTTC TTAGTTGAAA GGAAAGCGAT 2760

GATATCTGGG TGAAAGACGT TGAGGTAGAC AACACCAGCA CCTTGACGTT GCCCCAATTG 2820

GTTGGAGTAA GAGAAGCTGT CTTCAAAAAG CTTCATAACA GGAACGACAC CTGAAGCAGC 2880

TCCTTCATAG CCTTTGATAG GTGCACCAGC TTCACGAAGG TTGCTGAGGG TAATTCCCAC 2940

ACCACCACCA ATACGTGAAA GTTGAAGAGC TGAGTTGATA GAACGCCCGA TAGAGTTCAT 3000

ATCATCCGTC ACTTGGATTA GGAAACAAGA TACCAACTCC CCACGACGAG CACGTCCAGC 3060

ATTCAAGAAG GAAGGAGTAG CAGGTTGGTA GCGTTGGTGG ATGATTTCAT TGGCAATATC 3120

GATTGCAACA GCTTCATTCC CATCAGCGAA ATAAAGGGCA TTGAAGAAGA CACGGTCTTC 3180

CATATTTTCA AGATAGTATT CACCGTCATT AGTCTTTAAG GCATATTGAT TGTAAAATTT 3240

ATAAGCTGCC ATGAATGACT TGAATTGGAA GTTTTGGTCT TTGATAAATT GAGCTAATTC 3300

TTCCAAGAAC TCTGGACGGT ATTTCTTGAT AAAGGCTGTT TCGATGTAGT TGTGTTCAAT 3360

GAGGTAATTG ATTTTGTCTT TGATTGAATC AAAAACCATA GTGTTTGGAA CTACATTTTC 3420

TTTAAAGAAA GCATCCAAGG CTTCCTTGTC TTTATGAAGC ATGATTTGTC CATTAACAGG 3480

ACGGTTAATT TCGTTATTAA GACGGAAGTA AGTCACGTCT TCAAGATGTT TTAATCCCAT 3540

AAAATTTCCC TTATCTAATT ACAAAAGAAA GGCTTCTAAG TTAGCCCTAA AAGCAGTTTC 3600

TTCTGGATGA TGTACTAAGA TTATGCTAAT TGTTTCAGTT TTCCTGGTTG GAAACCTGAA 3660

AAGACTTCAG TTGGTGTTTG GATAACAGGA GCTGCGCTAA AACCGAGCTC TTTAACTTGA 3720

TCGACGTACT CAGGTTGCTC ATCAAGATTG ATTTCACGAT AAGAGACATT ATTACTGTCC 3780

AAGAAACGCT TGGTCATTTT ACATTGGACA CAATTGTTTT TAGAATAAAC GGTTACCATT 3840

GTGTAACTCC TCTTCAAAAT TTAATACTAT CTTAGTATAT CAGAAAATAA AATTTTGTCG 3900

GG 3902

(2) INFORMATION FOR SEQ ID NO: 213:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2456 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 213:

TATTGAAGCT ATTGTAGACT ACAAAGATAA GGATTTGCAG TTAGTAGGCG GTGAGACTCA 60

CTGATAACCT AAAAAGGATA GTCAATTATG CTTGTTTACT AACTATTAAC TATGCTAAAT 120

CAATTGAGGT TGTTTACATA AAACTCTATA TCAGAGAAGC CTGATATAGA GTTTTTTCTT 180

GCTAGTTTTA GGATTTTTTT GTAAAATAGA AAAAGTGAAG AGAGGTATGA AATGAGCAAG 240

AAAGATAAAA AAATCGAAAT TCAAGTAGCG GATGCCAAAG TTAATGTTGG TAAAGACAGT 300

TTTGAAGGTT ATACATTGAC TATCGGTAAA AAAGTTATCG GAGAAATTGC CGAATTAGAC 360

GGACAATTTG CCATTATAAA GAATGGGAAT GTCGATAGTT TTTATAAAAA ATTGGAAAAA 420

GCTGTGGAAA TTTTGATTGA AAATTATAAT TTAGCAAAAT AAGTCTTGTT TTTTTGAAAT 480

TTTCATGATA TAATAGTCCA TGTTGATTGT AGGAGAGATA GCGAAGAGGC TAAACGCGGC 540

GGACTGTAAA TCCGCCCCTT CGGGTTCGGG GGTTCGAATC CCTCTCTCTC CATTTCATTA 600

ATGGGGTATA GCCAAGCGGT AAGGCAAGGG ACTTTGACTC CCTCATGCGT TGGTTCGAAT 660

CCAGCTACCC CAGTTCTTAG GTAATAATCA AGATAGAAAG CAAAATATCT TAGGGTATTT 720

TATTTTTATA ATTGAAAGAC GTGAATGATA TGAACATGTC CTTGCGGGTG CTTAGGAAAA 780

AAATTATAAG TATGTCAAGT TTAAGAAAAA CTTGATTGTT GGAGGATTTT TTAGATGAAC 840

GAATTTGAAG ATTTGCTAAA TAGCGTTAGT CAAGTTGAGA CTGGTGATGT TGTTAGTGCT 900

GAAGTATTGA CAGTTGATGC GACTCAAGCT AACGTTGCAA TCTCTGGAAC TGGTGTTGAA 960

GGTGTCTTGA CTCTTCGCGA ATTGACAAAC GATCGTGATG CAGATATCAA TGACTTTGTT 1020

AAAGTAGGAG AAGTATTGGA TGTTCTTGTA CTTCGTCAAG TAGTTGGTAA AGATACTGAT 1080

ACAGTTACAT ACCTTGTATC TAAAAAACGC CTTGAAGCTC GCAAAGCATG GGACAAACTT 1140

GTTGGTCGCG AAGAAGAAGT TGTTACTGTT AAAGGAACGC GTGCCGTTAA AGGTGGACTT 1200

TCAGTAGAAT TTGAAGGTGT TCGTGGATTT ATCCCAGCTT CAATGTTGGA TACTCGTTTC 1260

GTACGTAACG CTGAGCGTTT TGTAGGTCAA GAATTTGATA CTAAAATCAA AGAAGTTAAC 1320

GCTAAAGAAA ACCGCTTCAT CCTTTCACGT CGTGAAGTTG TTGAAGCAGC TACTGCAGCA 1380

GCTCGCGCTG AAGTATTCGG TAAATTGGCT GTTGGTGATG TTGTAACTGG TAAAGTTGCT 1440

CGTATCACAA GCTTCGGCGC TTTCGTCGAC CTTGGTGGTG TTGACGGATT GGTTCACTTG 1500

ACTGAATTGT CACATGAACG TAATGTATCA CCAAAATCAG TTGTAACTGT TGGTGAAGAA 1560

ATTGAAGTGA AAATCCTTGA TCTTAACGAA GAAGAAGGAC GTGTATCACT TTCACTTAAA 1620

GCAACAGTAC CAGGACCATG GGATGGCGTT GAGCAAAAAT TGGCTAAAGG TGATGTAGTA 1680

GAAGGAACAG TTAAACGTTT GACTGACTTC GGTGCATTTG TTGAAGTATT GCCAGGTATC 1740

GATGGACTTG TTCACGTATC ACAAATTTCA CACAAACGGA TTGAAAATCC AAAAGAAGCT 1800

CTTAAAGTTG GTCAAGAAGT TCAAGTTAAA GTTCTTGAAG TTAACGCAGA TGCAGAACGC 1860

GTGTCACTTT CTATTAAAGC TCTTGAAGAA CGTCCAGCCC AAGAAGAAGG ACAAAAAGAA 1920

GAAAAACGTG CTGCTCGTCC ACGTCGTCCA AGACGTCAAG AAAAGCGTGA TTTCGAACTT 1980

CCAGAAACAC AAACAGGATT TTCAATGGCT GATTTGTTTG GTGATATCGA ACTTTAATCA 2040

AATTGAAAAT TCACAAAATC CTTTGTTTAC TAAACAAGGG ATTTTTCTGG CTCTTTGTCA 2100

ACTGTAGTGG GTTGAAGAAA AGCTAAGCTC GAGAAAGGAC AAATTTTGTC CTTTCTTTTT 2160

TGATATTCAG AGCGATAAAA ATCCGTTTTT TGAAGTTTTC AAAGTTCCGA AAACCAAAGG 2220

CATTGCGCTT GATAAGTTTG ATGAGATTAT TGGTCGCTTC CAGTTTGGCG TTAGAATAGT 2280

GTAGTTGAAG GGTGTTGACA AGCTTTTCTT TATCTTTGAG GAAGGTTTTA AAGACAGTCT 2340

GAAAAATAGG ATGAACCTGC TTAAGATTGT CCTCAATAAG TCCGAAAAAT TTCTCCGGTT 2400

CCTTATTCTG AAAGTGAAAC AGCAAGAGTT GATAGAGCTG ATAGTGGTGT TTCAGG 2456

(2) INFORMATION FOR SEQ ID NO: 214:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10974 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

AAATAGGATA TAGAGACATC CTTCTGATCT GCTTTTwACA AAGTCCAATT ATATGCGGAT 60

CTATACCTCC ACAATGTCCA TTATTATmCC TAACTATAAT ATGAGCCGAA AACACTATAT 120

CCTTAATGTC TCCATATCCA TCAGGGATAT TAATATTTAT TTTTCCACAA CTATATTGCA 180

TTGTAACCAT CTCCTTAAAC GACGCATTAT GATATTTGAT AGAGAAATTT TTATGAATAA 240

CTCAATAATT TTATAGTAAA TCATGCTTAT ATCTCAAAGA TACCTATTTT ATCTTGTCTC 300

GACCTTCTCC AAAGAATTGC TATAATACTA TTACAAATCC ATCTGCACTA CACTTCAAAT 360

TTTAGCACTG TATAAAAACG TTTCAATACA CTAACTTCAA GAAAACTTCC ACTATTAATT 420

GAAAAAATTG ATAGAGATAA ATTAAAAATC TATATTGAAA CTCATCCCGA TGCTTATTTG 480

ACTGAAATAG CTGCTGAATT CAACTGTCCT CCAACAACTA TTCATTACGC TCTAAAGGCT 540

ATGGGATATA GTCTAAAAAA GAGCCGTACC TACTGCGAAC AAGACCCAGA AAAAGTAAAT 600

CGGTTCCTTA AAGAATTGAA TCACTTAAGC TACCTGACTC CTATTTATAT TTATGAGACA 660

GGGGTTGAGA CCTATTTTTA TCTCGAATAT GATCGAGCCT TGAGCAGGCA GTTAGTCTCT 720

CTGGAAGAAG ATATAATTAT TTGAATTAAG ATCGAGACAA CGCACACCAG AGATTGCGAT 780

ACTGTTATAG AAGTACTAAT GCCCTTTTTT GTTTCAATAT ACTATGGCTC CGATGACCTA 840

TAAAGATACG ATGACGAGTG ACTTTTTCGA AGCTTGCTTC CAAAAATTCT TACTACCTAC 900

TTTAGATACA CCATCCCTTA TCATTATGGA CAATGCAAGG TTTCACAGAA TGAACATGTG 960

TAAGGAGCAG GGCATAGACT GTTACCACTT CCTACCTATT CACCCGAGTA TAATCCCATT 1020

GAGAAAATAT GGGCTTACAT CAAAAACATC TCAGAATAAT ATTGTCAAAT TACGATGCTT 1080

TTCTTGAGGC ACTTTTGTCC TATTCTTGTT TCAGCCGACT ATACTCCGTT ATTGGGCAGC 1140

TACGGAACAG TCGATGGGAC GATGGGGGGA CATAAAAAAA TCCTCCAGTT TTGTTTTTTA 1200

TAACAGTATA CTGGAGAATT GACAATCTCG GTAGATACCT CGTTATAGCG CGGTTACTTA 1260

TTAGGCAGTT ACAAAACAAC TGTGAACAGA AAACATTCCA GAGTCAGACA AGACTTTGGA 1320

ATGTTTTGGC TCTATAATTT CTGTAGTGGG TAATCCCACC CCAGGAATTA TAGGGTCGTT 1380

TCTTGTAGAA AAAAAGCCCC ATATGACCTA TAATGAAAAG CGTCTAACCA ACTCATTAGA 1440

AAGGGTTCAT ATGGAACAAC TTAAGAATAC CACAGATTTG CTCGGATTGG AAGACAAAAA 1500

TATCAAAATC TTGTCTGTTC TGAAATACCA AACCCATCTA GTCGTTCAGG CAAAGTTGGA 1560

TTCCCCCGCT CCTCCTTGTC CTCATTGTCA AGGGAAGATG ATCAAATACG ACTTCCAGAA 1620

AGCCTCTAAA ATTCCGCTTC TCGACTGTCA GGGTTTACCC ACGGTACTGC ATCTCAAAAA 1680

GCGCCGCTTT CAGTGCAAGA ATTGCCTTAA GGTGGTCGTT TCTCAAACAT CCATTGTCAA 1740

GAAAAATTGC CAGATTTCCA ACATGGTGAG ACAAAAAATC GCTCAGCTCC TCCTTGAAAA 1800

GCAGTCTATG ACTGAGATTG CCCACAGATT GGCGGTCTCA ACTTCCACCG TCATCCGAAA 1860

ACTGAGGGAA TTTAAGTTTG AAACCGATTG GACCAAGTTG CCAAAAGTTA TGAGTTGGGA 1920

TGAGTATAGC TTCAAAAAGA GCAAAATGAG CTTCATTGCC CAAGATTTTG AGTCCAAATC 1980

CATCCTCGCA ATTTTAGACG GGCGAACTCA TGCGGTGATT CGAAACCATT TCCAACGCTA 2040

TCAGAGAGAG GTTCGGGAGC TGGTCGAGGT CATCACCATG GACATGTACA GCCCTTATTA 2100

TCGGCTCGCT AAGCAACTCT TTCCAAAGGC GAAGATTGTT CTTGACCGCT TCCACATTGT 2160

CCAACATCTG AGCCGAGCTA TGAACCGAGT ACGAATCCAA ATCATGAACC AATTTGACCG 2220

AAAATCCTTG GAGTATCGGG CGCTCAAGCG CTTTTGGAAC CCTCGCTTTT TCGTTTCTAG 2280

GCTCGGGCTA AATCAGTCCA CTGGACTGAT TTACTACACC AGTATAGCTT CAAGCTCTGT 2340

CAGAAACGAT TCTATCAGCC CACGTTTCGA ATGCACTTAA CCCATCGGGA AGTACGAGAT 2400

AAGCTGCTTT CTTACTCTGA GGGATTACAG GTTCACTACG AACTCTATCA ACTCCTGCTC 2460

TTTCATTTTC AAGAGAAGAA TGCCGACCAT TTCTTTGGAT TGATTGAGCA AGAACTGCCA 2520

ACGGTTCATC CGCTTTTTCA AACGGTCTTT TGGACTTTTT TAAGGGATAG AGATAAGATT 2580

ATCAACGCAC TTAAGCTGCC TTATTCCAAC GCTAAACTTG AAGCGACCAA TAATTTGATT 2640

AAGATTATCA AGCGCAAAGC CTTTGGTTTC CGGAACTTTA ACAATTTTAA AAAACGGATT 2700

TTGATGACTT TGAACATCAA AAAAGAGAGT ACGAATTTCG TACTCTCCAG ATTGCAGCTT 2760

TTCGCCTACC CACTACACTT GACAAAGAGC CACTCTTTAT TCCATGGTAT CAAAGGCAAG 2820

ACTTGGTTTG GCATTGAGGT CCCAGCCTGC GAAGTTTTCT TTGTTCCACT CGCTGACGCT 2880

GGCATAGGCA ATCATACCTG CATTGTCTCC GCAGAGTCGC AGAGGGGGGA TGATAACCTT 2940

GACATCTGTG ATTTCGGCTG CTAGGCGTTC TCTGAGACCT TTATTGGCTG CCACACCACC 3000

TGCCACAACT AGGATTTTAA CAGGATATTT CTCCAAAGCC TTCTTGGTTT TTGCCATGAG 3060

AATGTCCATA ACTGCTGCTT GGAAGGAAGC ACACAAATCT TCTGTAGACA GGCTTTCTCC 3120

CTTTTGCTCG GCATTGTGAT GAAGATTGAT AAAGGCAGAT TTCAAACCTG AGAAGGAGAA 3180

CTCCAGATTA TCTTCCTTAA TCATGGCACG GGGGAAATCA TAAATATCCT GCCCCTGATG 3240

AGCCAGCTCG TCAATCTCAC GACCTGCAGG ATAGGTCAAG CCCATGACAC GGCCGACCTT 3300

ATCATAAGCC TCACCAACCG CATCATCACG GGTTTCCCCA ACAATCTTAT AATCTCCTGC 3360

CTCCGAAACA TAAACCAACT CTGTGTGTCC GCCGCTGACC AAGAGGGCTA GCAAGGGAAA 3420

CTCCAAAGGC TCCACACTCT GAGCTGCCAT GAGGTGCCCA GCCATGTGAT TAACAGGAAT 3480

CAGTGGAAGT CCGTGAGCCC AAGCAAAGGC CTTGGCAGCT GACAAACCAA CTAGCAAGGC 3540

TCCGACCAAG CCTGGTCCGT AGGTAACCGC AACAGCTGTC ACGTCCTCTT CGGTAATCCC 3600

TGCTTCTGCC AATGCCTCCT CGATACAGGC TGTAATGACC TCGACATGGT GACGACTGGC 3660

TACTTCGGGC ACTACGCCAC CAAAACGTTT GTGACTCTCA ATTTGACTAG CAATGACATT 3720

GGACAAGAGC TCATCGTCGT TTTTCAAGAC GGCGACACTG GTCTCATCAC AGGATGTCTC 3780

AAATGCTAAA ATATATCTAT CCTTCATCTA TTTCTCTCTT CATGATAATG GCGTCCTCGA 3840

CTGGGTCATG GTAGTAGGCC TTTCGCTCAG CGATAACTGT CATCTTTTCT TTCTTGTAAA 3900

ATGCTTGCGC TCGTTGATTT GACTGTCTGA CTTCGAGGAA AATTTCCTTG TCTGTCGGCA 3960

ATTGAGCAAA CAAGGcTGAC GCAATCCCCT GACCCTGATA AGCTCCTTTG ACAGCGATTT 4020

GCAGGACTTC TGCTTCAAAA AGATTCTCCT GCACAGCTAG AAATCCAATC ACTTCTGCCC 4080

CATCATAAGC CAATGCATAC CAAGTCTGGT CTTGGGACAG ATCTGCTTGG ATTTGCTCCA 4140

GAGTCCAAGG ACTGACTAGG TAAACAGCTG CCATAACAGC GTAGATGGCT TGAGCTAGGT 4200

CAGGCTGTTG TTGAATTCGC TTGATTTCTA TCATAGGCGT TTAATGTAAG ACTCGCCAGA 4260

CTCGGTATGG TTCTTGAGCC AGTTTTCCTC AGCCTCGACT CGTTTGAGGT AATTCGGCAC 4320

AAAATCATGC AAGGAGTCTG CTTCCTTGTC CCAGGCCAAA AGAGCTAGAT TAGCTGCATT 4380

GGGCAATGTT TCTTTGTAAT CAGTCCTTGG CAAGTGTTTT TGAATCTGCT CAACAAAGGG 4440

GCCAACTTCT CCGACAAAGG TTACCTGACT AGTACCCTTG ACTTTTTCTA GCACCTCTTC 4500

AAAAGATAGG TGCGCTTCTG CCATGACAGG TTTGGCATTT TCATAAAATC CTGCATAAAC 4560

ATTATTGCGA CGCGCATCCA TCAAGGGGAC AAACAAACCT TCTTGTTGAT GGGGCACCAG 4620

AGCCAAGAGA CTCGACATAC CAACCAACTC GATGTTCAGG GTGTGAGCTA AGGTCTTAGC 4680

AGTTGCTACC GCAATTCGCA AGCCTGTATA GCTACCCGGC CCTTCAGCTA CCACGATTCG 4740

GTCCAAATCC TTGGGTGTCC AATCCAAACT TGCCATCAAA AAATCGATGG CAGGCATAAG 4800

AGTAATACTG TGATTTTTCT TAATATTAAT CGTCGTCTCG GCAAGAACCT GCTTATCCTC 4860

TAAAATAGCC AGAGAAAGAG CCTTGCTGGA CGTATCAAAA GCTAATACTT TCATAACACA 4920

TTCCTATCTT TTTGTCTGCT TACTATTATA CTACAAAAGC TGGCACATGG GAATTTTCTT 4980

TGCCCCCAGA CAAGAGTGCC CTCACTTAAC TAAAAATAAT TTAAAAAAAT GCTCACTTTT 5040

CCTTTTCTTT TCCGAATATA AAAGTGAACA AGAAAAAAGG AGGAAAGTTC AATGACAAAT 5100

TTTGACATTC TTGACAATCA ATTTTTATCC TTATCTGAAA ATGAATTATC AGATATTGAT 5160

GGCGGTCTCG CTCCCTTGGT TATCTTTGGA GTAGCAGTAT CTTGGAAGGC TATTGCAGGT 5220

GGAACAGCAC TTATAGGTTC TGGTTTGGCA GCTGGTTATT TTTTAGGAGG AGATTAATAT 5280

GATGAAAGAT TTGAACAATT ATCGTGAAAT TTCTAATAAG GAATTGCAAG AAATCAAGGG 5340

TGGCTTTGGT GTCGGTGTTG GTATCGCTTT ATTTATGGCA GGTTATACCA TTGGAAAAGA 5400

CCTTCGTAAA AAGTTTGGTA AGTCATGCTA GATAAGAAAC ACATTTTTAG AAGGATAAAT 5460

TTTATTGTCT TCATCTCTTA CAGTTTGCTC AGCATTCTCA ATGATTTGAA CATTACTACC 5520

ATCCCTTTAC CATTCGATTT ATCTGTTTGT ATTGTTTTAT TTTTATGCTT CAACTCTATT 5580

TTTGATCAGA ACAATGACTC CCATAAAAAT AATAAGCTTT GAAAATTCCA TTGTCATGTC 5640

ATGTTAGAAA AATGCAAAGA CCACCTCATC TTGATAGATG GGGTGGAATT TTCGTGTCGT 5700

AAATCTACTA TCTCTACATT CCCAAACAAA AAACCCCAGC ATAAGCAGGG CATCTAAGCA 5760

TTTAATTCAA AGTAAAATAC AAACCAAACG ACATAGGTCA CGAGGAGGAG AAAAAGCGAG 5820

TAGAGAGTCA CAAAGGTCAT TTTCCACAAG AACTTGGTTT GTCGTCGTTC CAGTTTGGCA 5880

AATAGAAGAT TCCCCGCATA AACGCAAGCA ACAAAAACAA TAAAAGCTAC CAAGCGAGCT 5940

CCGATAGCAA AAGCAAATAA GTTATACATA GGGCAACCTC CTTGACTTAA AATCTATATG 6000

GAATTATGAC AAGCAATAAA TTTCACTTCC GTTATCAACA TAATACATTT TCTTTATTTT 6060

TGAAAACGCT TACCAAAGAA ATCGTCCCCT AACTTTCTCG TTTCCGTCTT TTACTAATTT 6120

TTCATTTTGT GGTATAATTG AAATAATTGT AACGAATCAA GGTCAATCTA GACACAAAAT 6180

GGAATGAAAT CAAGCAAATA TCTGCTAAAA GTTTGGAATA AGCTGACCTG TAAATAGAAA 6240

GGAACTATAT GATTTACAAA GTTTTTTATC AAGAAACAAA AGAACGTAGC CCACGCCGTG 6300

AAACAACACG CACGCTTTAC CTAGACATCG ATGCCAGCTC AGAACTTGAG GGCCGTATCA 6360

CTGCTCGCCA ACTTGTCGAA GAAAATCGCC CAGAGTACAA TATCGAGTAT ATCGAACTCT 6420

TGTCTGACAA ATTGCTCGAT TACGAAAAAG AAACTGGCGC CTTCGAAATT ACGGAGTTCT 6480

AATATGGCCT ACACTCTTAA ACCTGAAGAA GTCGGCGTTT TTGCCATCGG TGGTCTAGGA 6540

GAAATCGGGA AAAACACTTA CGGAATTGAA TACCAAGACG AGATTATCAT CGTCGATGCT 6600

GGGATTAAAT TCCCAGAAGA TGACTTGCTT GGTATCGACT ATGTCATTCC TGACTACTCT 6660

TACATCGTGG ACAATATCGA CCGCGTCAAG GCTGTTTTAA TCACACACGG ACACGAGGAC 6720

CACATTGGTG GGATTCCGTT CCTACTCAAG CAAGCAAATG TCCCTATTTA TGCTGGACCG 6780

CTTGCCTTGG CTTTGATCCG TGGGAAACTC GAAGAACACG GCCTCTTGCG CAACGCCAAA 6840

CTTTACGAAA TCAACCACAA CACCGAGTTG ACCTTTAAAA ATCTCAAGGC AACTTTCTTT 6900

AGAACGACTC ACTCTATTCC AGAGCCTTTG GGGATTGTCA TTCATACTCC TCAAGGGAAA 6960

ATCGTCTGTA CGGGTGACTT TAAGTTCGAC TTTACTCCAG TTGGAGAACC TGCGGACTTG 7020

CATCGTATGG CTGCGCTTGG TGAAGAAGGC GTGCTCTGTC TCCTGTCTGA CTCGACAAAT 7080

GCGGAAGTAC CAACCTTTAC CAACTCTGAA AAAGTCGTTG GTCAGTCCAT TATGAAGATT 7140

ATCCAAGGTA TTGAAGGACG TATCATCTTT GCATCCTTTG CCTCAAATAT CTTCCGTCTC 7200

CAGCAGGCAA CAGAAGCTGC TGTTAAGACT GGACGCAAGA TTGCGGTCTT TGGTCGTTCT 7260

ATGGAAAAGG CCATTGTCAA CGGAATCGAT CTTGGCTACA TCAAAGCTCC TAAGGGAACC 7320

TTTATCGAGC CAAATGAAAT CAAAGATTAT CCTGCAGGAG AAGTTCTTAT CCTCTGTACA 7380

GGTAGTCAGG GTGAGCCTAT GGCAGCCCTC TCTCGTATCG CCAACGGAAC CCACCGTCAA 7440

GTACAATTAC AACCAGGTGA TACCGTTATC TTCTCTTCTA GTCCCATCCC TGGAAACACT 7500

ACTAGTGTCA ACAAGCTGAT TAACATCATT TCTGAAGCTG GTGTCGAAGT TATCCACGGT 7560

AAAGTGAACA ATATCCATAC ATCTGGACAC GGTGGTCAGC AAGAGCAAAA ACTCATGCTC 7620

TGCTTGATTA AGCCAAAATA CTTCATGCCT GTCCACGGTG AATACCGCAT GCAAAAAGTC 7680

CACGCTGGAC TAGCAGTGGA TACTGGTGTT GAGAAGGACA ATATCTTTAT CATGAGCAAT 7740

GGCGATGTGC TTGCCCTTAC TGCTGACTCA GCTCGTATCG CAGGTCATTT CAACGCCCAA 7800

GATATCTATG TCGATGGAAA TCGTATCGGT GAAATTGGCG CAGCTGTCCT CAAAGATCGT 7860

CGCGATCTAT CTGAAGACGG TGTCGTTCTG GCAGTTGCAA CTGTTGACTT CAAATCGCAG 7920

ATGATTCTAT CTGGTCCAGA CATCCTCAGC CGAGGCTTTG TCTACATGAG AGAGTCTGGC 7980

GACTTGATTC GCCAAAGCCA GCGTATCCTC TTCAATGCCA TTCGTATCGC ACTGAAAAAT 8040

AAGGATGCTA GCGTGCAATC TGTCAATGGT GCCATTGTCA ACGCTATTCG CCCCTTCCTC 8100

TATGAAAATA CCGAACGTGA ACCGATCATC ATCCCGATGA TCCTCACACC AGATGAAGAA 8160

TAAAGCAAGA AAACAGCCCC GTCCTCGGAG CTGTTTTTCT CTATGCTTTC TTTTGAGATT 8220

AAAACTCATA CTCAATGAAA ATCAAAGAGC AAACTAGGAA GCTAGCCGTA GGTTGCTCAA 8280

AGCACTGCTT TGAGGTTGTA GATAGAACTG ACGAAGTCAG TAGCCATACC TACGGCAAGG 8340

CGACGTTGAC GCGGTTTGAA GAGATTTTCG AAGAGTATCA ATAAAAATCG AAATCAGACT 8400

AGAAGGCTAA GCGAAAGCAT AACTTGAGTT AGCTCCCATA GTTCGGGAAA CTATGGGAGG 8460

CTGGAGATGA ATCAAAGCCA AGCTTTGAAC TCATTCGTAA GAAGCCGACG ACGTATCATT 8520

TTGATTTTTG AAGAGTTTTA GAAATACTAC GATTTTTACC TTCCAGATAC ACCATCAAAA 8580

TAGAAATATC TGCTGGGTTT ACTCCCGAAA TACGGCTGGC TTGGCCGATG GTTTCTGGAT 8640

TGATGAGTTT GAACTTCTGA CGGGCTTCGG TTGCGATAGA ATCAATGTCA TCCCAGTCGA 8700

TATTGGCCGG AATGCGTTTT TCTTCCATGC GTTTCATCTT GGCAACCTGG TCCATGGCTT 8760

TGGAAATATA GCCTTCATAC TTGATTTCTG TTTCAATCAA TTCGATAATC TTGTCATCCA 8820

AGTCTTCTGC AGCTGGTCCG ATGAAGGCCA CCACATCTTG GTAAGAAACT TCTGGACGGC 8880

GAAGGAATTC CTTGGCTGTC ACTGCATCGG TCAAGGGTTT GAAGCCCATC TCCTCAACCT 8940

TGGCATTGGT TTCCTTGACT GGCTTGAGTT TGATACTGTC TAGGCGCTTC ATCTCATTAT 9000

CAAATTGATT TTTCTTGATT TCAAAACGAG CCCAGCGTTC ATCGTCCACA AGGCCAATCT 9060

CGCGTCCCAT CTCAGTCAAG CGCATATCAG CATTGTCATG ACGAAGAATG AGACGGTATT 9120

CAGCACGACT GGTCAAGAGA CGGTAGGGTT CAATGGTTCC CTTGGTCACC AAGTCGTCGA 9180

TCATCACCCC GATATAACCA TCACTGCGCT TCAAAATCAA TTCAGGCTTG CCTTGGATTT 9240

TCAGAGCCGC ATTGATACCC GCGATAATCC CTTGGCCTGC TGCCTCTTCG TAACCTGATG 9300

TTCCATTTGT CTGACCAGCA GTGAAGAGAC CTGAGATTTT CTTGGTTTCC AAAGTCGCAC 9360

GCAACTGATG AGGCAAGACC ATATCATACT CAATAGCATA ACCTGTCCGC ATCATCTCTG 9420

CATTTTCCAA ACCTTTGATG GAATGCACCA AGTCACGCTG GACATCCTCA GGCAGACTGG 9480

TTGAAAGTCC TTGCACATAG ACTTCCTCAG TATTGCGCCC TTCTGGCTCA AGGAAGAGTT 9540

GGTGACGTTC CTTGTCCGCA AAGCGCACAA TCTTGTCTTC AATCGACGGA CAGTAACGAG 9600

GCCCCACTCC CTTGACCACA CCTGTAAACA TAGGCGCACG GTGGAGGTTG TTTTGGATAA 9660

TCTCATGACT GGTACCATTG GTATAGGTCA ACCAGCATGG TACTTGGTCC TTGACATAAT 9720

CCTCATCACG TGAAGTGTAT GAGAAATGAT TAGGCACTTC GTCTCCTGGC TGAATTTCTG 9780

TCACATCGTA ATTGATAGAA GAAGCCTTGA CACGTGGAGG GGTTCCTGTC TTGAAACGAC 9840

CGATTTCGAG ACCCAGTTCC TTGAGATTGT CAGCTAGGTT AATAGAAGCC AAGCTGTGGT 9900

TAGGACCTGA TGAGTACTTG AGGTCTCCGA TGATAATTTC CCCACGGAGA GCAGTCCCTG 9960

TCGTCACAAT AACAGCCTTA GCAGCATATT CTTGATGGGT GGCTGTACGC ACACCGACAA 10020

CCTTGCCATC TTCCACCAAA ATCTCATCAA TCATGGTTTG ACGAAGGGTC AGATTTTCTT 10080

GGTTTTCAAC CGTCTTGCGC ATCTCCTTAG AGTAAAGTTC CTTGTCAGCC TGCGCACGAA 10140

GGGCACGGAC AGCTGGCCCC TTCCCTGTGT TTAGCATCTT CATCTGGATG TAAGTCTTGT 10200

CAATGGTTTT GGCCATCTCG CCACCGAGGG CATCGACTTC ACGCACGACA ATCCCCTTGG 10260

CAGAACCACC GATAGAGGGA TTACAAGGCA TGAAAGCCAG CATTTCAATA TTGATGGTCG 10320

CAAGCAGGAC CTTACAGCCC ATACGGCTAG CGGCCAAGGA AGCCTCAACC CCAGCGTGTC 10380

CCGCACCAAT TACAATAATA TCGTATTCTT CAGTAAAATG ATAAGTCATG TTTCTCTCCT 10440

ATTCCTCAAG ATGAATGTGT CTTAGTTGGC CTTCCCAATC TGGTAGGGCT GTTTTTAAAA 10500

AGACTGGAAC TAGCTGGATA TTCTGGAGCT TATCCAAGTC AATCCACTCA CAGGGCTGCC 10560

TTTTCTCATC TTCCTGCATG GTCAACGGGG CATCTTCAAG CAAATCCACC AGATAATGAA 10620

ACTCGATATT GTGATAGGAA ACGCCGTCCA CTTCAAAACG ATTTTCAACC ACAAAAGCTA 10680

GCTGCCCAGC TTGAGCTTTG ACACCCAGTT CTTCCTTCAC TTCACGGACT ACCGCGTCTT 10740

CCGTGCTTTC ATTGACTTGA ATCGCACCTC CAATAGTGTA ATACTTGCCC TTGTCTTTGG 10800

TAACTAGAAG CTTGTGATTT TGGACAATCA AGGCTGTAGC CCGAACACCA AAAACCGTAT 10860

TGTCTACTTT TGTCCGAAAG TCTTGTTGAG TCATTCTTGT CCTTTCCCTT AAACGACACA 10920

AAAACAGTCA AAACTACAAA GAAGTGCAGG ACAAAAAAGC CTGCAACATC CAGG 10974

(2) INFORMATION FOR SEQ ID NO: 215:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 987 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

CCCGTTATGA TTATGGATAG CGCTTTCAAA TTTTTAAACT CCTATCCCAT CCTTTTATCT 60

ATATAATAAG TGAAAATATA ATAACTGTCA AGTAACTGAA GTGAATTTTA TAAAAAAATT 120

1220

ACAAGCCAAA TTTGTAAAGT TTACACTAAG CCGCTAGgCA ATCGTCTATC AGAATATCCG 180

TTTATTTGTC AATAATCCGA GAAAATCTTG CAACGCTTAG AAGTCTATAA AAACTATCAA 240

CATTTATATG ACTTGCGAAT AGCAATCCTG CTAAACCTTT CCACACTCTA TCTATACAAT 300

CAAGATAAAA ACATGTGTAA GCAAATCTGC TACACTTTAC TGGAGGACGC CAAGAATAAG 360

AAAAGCTACG ATAGGCTTGC TATCTGCTAT GTCCGTATTG GGATTTGTAC AGACGATTCT 420

AAACTTATCC AAAAAGGGTT CTCCCTTCTG GAGCTGACCG AGGAAACTTC TATGCTGTCT 480

CATCTCAAAA AAGAAGTAGA GACCCATTAT CAACCAAAGA AATTATAAAA AAAGTCGAGG 540

GAGCTCCTCG ACCTTTTCAT AGAATCGCCG AACGATTTAA CGAGAAAGTA TGACTTTTAC 600

GTTTATCCCA ACTCAATTAT GACATTTTTT TCAAAAGTCA ATATATCTCA CTTTTTCAAC 660

GACAAGAAAG AGGCTGATAA TCTACCAACC TCTTATTCTG AACCCATCAC TCCATCACTT 720

TTTAGCTTCA TTCGCTTTCT TAGCGACTGC AATCTGGTAT TCGACTTGGT CATTCCCCTT 780

ACCGGTACAA CCATGAGCAA TTGTAGTCGC TCCTATCTGA TGCGCTATTT CAACCAATTT 840

TTTAGAAATC AGAGGGCGGC TCAAGGCAGA TACCAAGAGA TACTTTTGTT CATAATAGGC 900

ATGTGACTGA TGAGCCACTA GCACATAATC TGTAGCAAAT TCGTCCTTAA CATCAATGAC 960

ATAAGATTCT ACTGCCCAAA CCTTAAG 987

(2) INFORMATION FOR SEQ ID NO: 216:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2651 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 216:

CTGGGTCTTG TTCATAGTAG GTGTGGTtCT TTTTTTCGAG TGTAGCCCAT AGCTTTGAGC 60

GCATAGTGGA TGGTAGTTGG ATGACAGCCA AAGTCAGAAG CTATTTCAGT CAAATAAGCA 120

TCTGGATTGT CAGTAAGATA GTTTTTAAGT CTATCTCTAT CAACTTTTCT TGGTTTTGTT 180

CCTTTTACTT GGTGGTTTAG CTCTCCTGTT TTCTCTTTTA GCTTTAACCA GCCATAAATG 240

GTATTACGTG AGATTTGGAA AACGTGTGAT GCTTCTGTTA TACTACCTAT TCGCTCACAA 300

TAAGAGAGAA CTTTTTTACG AAAATCTATT GAATATGCCA TAAGAAGATT ATACCACATT 360

GTGTACTATT TTTGGTTCAT TTTACTATAT TTTATAAGTT ATAGTGTAGC ATTCCAACTT 420

CAAAGCACTA TAAAGTAAAT TGAAACAAGA ACAATACAAA CAATTCTCGT AAACGGATTG 480

CAACCACAAA AAAGCAAGCA TTCACAAGAA TACTTACCTA TCATGGGAGG AACAACCGTT 540

CCTCTTTTTT ATTACTAAAA TTCAAAGAAT TCCAATGCTT TTTTCAAGAG CAAATCCGTA 600

TATTCTGGAT CTTCTTGGGC TACTTCTATT TCCCGCTGAA CTTTTTCCAA ATCATCTGTA 660

ATCACTCCAT CTACTCCTAA GTGAAGAGAT TTGCTGATAG CTTCTGAATC ATTGACAGTC 720

CAGACATAAA GTTTCTGATC CGTTGTCCAT AGTTTGCTTA CAAAATATTC ATCCAAGGTT 780

GAGTACTCCA TAGTATATCC TGTCGCTCTT GTTTTAGGAA AGACAGAATT GTAGGGCATG 840

ATGAAATAAA CTGGTAGTTC GGCATCATAC TGTCTTACTT TTTCGACAAC ATGGTAGTCT 900

AAAGACTGGA TTTGATGTCC ATAAATCTTG AGCTTTGCAG CATAACGGGC TAAAAAGCGG 960

TTCATCATGT CTGGACTATC TTTTTTACTG GTTTTAATTT CAATTAGTAA TTTTTGACCA 1020

AGTTCGTTGG CTCGACTGAG ATAATCTTCA AAGCTTGAAA TTTTAGTCTG GTAGCCATTT 1080

TCAAAAATAT CAATCCCTTT AAGCTCCTCC AAGTTTAAGT CTTGAGGACT TTTATTGATA 1140

CCTGCTAGAT TTTTCAAGTT AGCATCATGC ATCATGACAA ACTGCCCATC TTTTGTTTCC 1200

TGCACGTCCG TCTCCACCAA GTCTGGTTTG AGTTGTGCTG TAGTTTCCAA GGACTCTACT 1260

GTATTTTGAA TCCCATTTGC ATTGGAAACC CCTCGGTGAG AAATAAGTTG AGGTAGATGA 1320

ACCATGGGAG CCTCCAGATA AATATAACCT TCTAAGGCAA AGAAAAGACT GGCACAAGTC 1380

ATGACACCCC ATCGCACGAT GTGATCTTTT TCTCTCCTAG GAAGCATATC CAGCTCCTTT 1440

CCTGTCAAAA ATGAAACAAA TTTAACCAAA AAATAAGTCA GAGCCATATA ATAGAGATTT 1500

TTAATCACGA CAAAATTCAA AATACCAAGA ATCAGAGACT CTCTCTGAGT GATATCATCT 1560

ACCAAAGTTT GAGCCAATAA TAAAGGAATC AAAGGAAGAT ArAATAATAA ATGTGCTTTG 1620

AGCAAGATGT AAAATAAATT CCAAGCATAA AAAGTAACTC TCTTCTTGGT TTTCTCCAAG 1680

CTAAACATCA CTGCTTCTCG AACAGTCAGC TGATCATATA CAATCTTCGG AAGGGCAAAC 1740

ATCAATCTGA CAGAGACATA GAGAAAGATA AGAGATAGAA GTAGGATGCT CAGCCACCAC 1800

ATCCAATATC TATCTTCTAA ATAAGCTTGG ATAAACTCTG GAATGACGAT TTTATTAAGA 1860

TAATAAATCT TCAGCATTTT CCGTATAAAA GGAAACAGCA TAGCTATATA GAAAAAGATA 1920

AACAAGGCTT TAGCGCAAGT TAGCTTTTTC ATAAATCCAA AACTTTCATG GAAAACCTTG 1980

CGGATATACT CAATTAGCCT TCGCTTTTCA TTATAGAGGA GATGACGAGC ACCAATAAAG 2040

AGGAGTCCTA TTTGAAAATA AGCAACCAGA AGGTTAATTA CAATCAAGGC TAAAAAAGCT 2100

AGACTAATCA ATGGAGAATG AGTAAGGATG GCTAAGACAT TGTTATAGGA AATAAAAAGA 2160

TAACCTGTCT GATCTAATAA GAAGCTAGCC AACCATGAAT TGAATGGTAC CCACAAATAC 2220

TCCACTATCA TAAAAATCAA GAAAAATAGA AAGAGGATTT TATCAAGATC GAGGTAAATC 2280

TGTTTAAGAC CCAATTTTTT AGGTTTTTCA GGTTTCATAG GCACTCCTAG TCAAATAATT 2340

GAGACAAGTC CAAGCCACCA AAAGGATTGT TTGATAAGCT ACTTTCTGTC TCTAACAATT 2400

CCCTAGCTTG ATCCGACTCT AAGAAGGATT CGTAAACACG CGCCGTCATC CGAGCATCCT 2460

CTAAACTATT ATGAGACTGA CCTTGAAATC CAAGAAATGA GGCAACAGTT TGCAATTTGA 2520

GATTGGCAAT ACCATGTAAA TCTGAACTCC GACGTTCAAA AGCTTCATCA TACAAATCCA 2580

CCTTGTACTG TTGGCTATAG TCTAAACCAT GCTCTGCTAA AATAGGTAAA TCACTTTTAG 2640

CAGCATTGTA G 2651

(2) INFORMATION FOR SEQ ID NO: 217:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5638 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 217:

CGTTATAATA AACTTGTGAA AAAATTAACA AAGGATATCG TTCCTTGAAA GCTATGGAGG 60

AAAATATGGC TGATAAAAAA ACTGTGACAC CAGAGGAAAA GAAACTCGTT GCTGAAAAAC 120

ACGTAGATGA GTTGGTTCAA AAAGCTCTAG TTGCCCTTGA AGAAATGCGT AAATTGGATC 180

AAGAACAAGT TGACTACATC GTTGCCAAAG CATCAGTAGC AGCTTTGGAT GCCCACGGAG 240

AATTGGCTTT ACATGCCTTT GAAGAAACAG GACGTGGTGT ATTTGAAGAC AAAGCAACTA 300

AGAACTTGTT TGCCTGTGAA CACGTAGTAA ACAACATGCG CCACACTAAG ACAGTTGGCG 360

TTATCGAAGA AGACGATGTA ACAGGATTGA CTCTTATTGC TGAACCAGTT GGTGTTGTTT 420

GTGGTATTAC TCCAACAACA AACCCAACAT CAACAGCAAT CTTCAAATCA TTGATTTCAT 480

TGAAGACACG TAACCCAATC GTCTTTGCCT TCCATCCATC AGCACAAGAA TCATCTGCTC 540

ATGCAGCTCG TATCGTCCGC GATGCAGCTA TCGCAGCTGG TGCTCCTGAA AACTGTGTGC 600

AATGGATTAC TCAACCATCT ATGGAAGCAA CAAGTGCCCT TATGAACCAC GAAGGTGTTG 660

CGACAATCCT TGCAACAGGT GGTAATGCCA TGGTTAAGGC GGCTTATTCA TGTGGTAAAC 720

CAGCTCTTGG GGTAGGTGCC GGAAACGTTC CAGCTTATGT TGAAAAATCA GCAAACATTC 780

GTCAAGCAGC ACACGATATC GTCATGTCTA AATCATTTGA TAACGGTATG GTCTGTGCAT 840

CTGAACAAGC AGTTATCATT GATAAAGAAA TTTACGATGA ATTTGTAGCA GAGTTCAAAT 900

CTTACCACAC TTACTTTGTA AACAAAAAAG AAAAAGCTCT TCTTGAAGAG TTCTGCTTCG 960

GCGTCAAAGC AAACAGCAAA AACTGTGCTG GTGCAAAATT GAACGCTGAC ATCGTTGGTA 1020

AACCAGCAAC TTGGATTGCA GAACAAGCAG GATTTACAGT TCCAGAAGGA ACAAACATTC 1080

TTGCTGCAGA ATGTAAAGAA GTTGGCGAAA ATGAGCCATT GACTCGTGAA AAATTGTCAC 1140

CAGTTATTGC AGTTTTGAAA TCTGAAAGCC GTGAAGATGG TATTACTAAG GCTCGTCAAA 1200

TGGTTGAATT TAACGGTCTT GGACACTCAG CAGCTATCCA CACAGCTGAC GAAGAATTGA 1260

CTAAAGAATT TGGTAAAGCT GTTAAAGCTA TTCGTGTTAT CTGTAACTCA CCTTCTACTT 1320

TTGGTGGTAT CGGGGACGTT TACAATGCCT TCTTGCCATC ATTGACACTT GGATGTGGTT 1380

CTTACGGACG CAACTCAGTT GGGGATAACG TTAGTGCCAT TAACCTCTTG AATATCAAAA 1440

AAGTCGGAAG ACGGAGAAAT AACATGCAAT GGATGAAACT TCCTTCAAAA ACATACTTTG 1500

AACGTGATTC AATTCAATAC CTTCAAAAAT GTCGTGACGT TGAACGTGTC ATGATCGTTA 1560

CTGACCATGC CATGGTAGAG CTTGGTTTCC TTGATCGTAT CATCGAACAA CTGGACCTTC 1620

GTCGCAATAA GGTTGTTTAC CAAATCTTTG CGGATGTAGA ACCGGATCCA GATATCACAA 1680

CTGTAAACCG TGGTACTGAG ATTATGCGTG CCTTCAAACC AGATACCATC ATCGCACTCG 1740

GTGGTGGGTC TCCAATGGAT GCTGCCAAAG TAATGTGGCT CTTCTACGAG CAACCAGAAG 1800

TGGACTTCCG TGACCTTGTC CAAAAATTCA TGGATATCCG TAAACGTGCC TTCAAGTTCC 1860

CATTGCTTGG TAAGAAGACT AAATTCATCG CGATTCCAAC TACATCTGGT ACAGGATCTG 1920

AAGTAACACC ATTTGCCGTT ATCTCTGATA AAGCAAACAA CCGTAAATAC CCAATCGCTG 1980

ACTACTCATT GACACCAACT GTGGCAATCG TAGATCCTGC TTTGGTATTG ACAGTTCCAG 2040

GATTTGTTGC TGCTGATACT GGTATGGACG TATTGACTCA CGCGACAGAA GCATACGTAT 2100

CACAAATGGC TAGTGACTAC ACTGATGGTT TAGCACTTCA AGCCATTAAA TTGGTCTTTG 2160

AAAATCTCGA AAGCTCAGTT AAGAATGCAG ACTTCCACTC ACGTGAGAAA ATGCATAACG 2220

CTTCAACAAT CGCTGGTATG GCCTTTGCCA ATGCCTTCCT AGGTATTTCT CACTCAATGG 2280

CCCATAAGAT TGGTGCGCAA TTCCACACAA TCCACGGTCG TACAAATGCT ATCTTGCTTC 2340

CATACGTTAT CCGTTACAAC GGTACACGTC CAGCTAAGAC AGCAACATGG CCTAAGTACA 2400

ACTACTACCG TGCAGATGAA AAATACCAAG ATATCGCACG CATGCTTGGA CTTCCAGCTT 2460

CTACTCCAGA AGAAGGGGTT GAATCTTACG CAAAAGCTGT CTACGAACTC GGTGAACGTA 2520

TTGGGATCCA AATGAATTTT AGAGACCAAG GAATTGACGA AAAAGAATGG AAAGAACATT 2580

CTCGTAAATT AGCCTTCCTG GCTTATGAAG ACCAATGTTC ACCAGCTAAC CCACGTCTTC 2640

CAATGGTAGA CCATATGCAA GAAATCATCG AAGATGCATA CTATGGCTAC AAAGAAAGAC 2700

CAGGACGCCG TAAATAATTG TTTATCAGTC TAGAAGCAAG ACAAAAACTC AATTTGAGGG 2760

AAAGATCCAG TAATTTTTCT ATGATAAAAG GCATCCTATC AAGGTTTTTG AACACCTGAT 2820

AGGATGCCTT TTTATGATAT TGAGGCCTTT TTGCCCTTTT TGAAAAACTA GAATAGAAAC 2880

AAAATATATA ATAGATTGAA ACTAGAATAG TACATATCTG CTTCTAAAAC ATTGTTAGAA 2940

TTCGATTTGA CTGTCCTGAT CGATTTGTCC TGTTCTTATT TCATTTTGAT ATATAAAAAA 3000

TATAGTATAG TAGACTGAAT CTAAAATAGT ACGAAACAAT TGCTAAAACA TTTATAGAAA 3060

TTAATTTTAC TTTTCTGATA GAGTTGTTCA CATCTTATTT CAATTCACTA TAGTTTAATT 3120

TAAGAGTAGT ATTTACTAAG GCCCAATTAA AATCAAAGAG CAAACTAGAA AACGAGTGCC 3180

ATTCAGCTCA AAACACTGAT TTGAGATTGC AGATAAGACT AGCCCCCTCA TTAACAGATT 3240

TACGATAAAA CGATGACAAG GTGTGTTGCT TTTTGATTTC TAAAGAGTAT AATGATAGAT 3300

CTCTATAAAA TAAGTGCGAA GGAAATGAGC TTTTATAGTC CTTTCGTTTT AAAATACTAT 3360

CTCAGATATT CTTATATCGA CAAGAAGTTT TTGAGTCATT CCCTCATCAT ACATATTAAA 3420

TAAATAGTGG CTCATTCAAT TTTTCACTAG AATAATAAGC TAGTATAGTA AACTGAAATA 3480

AGATATAAAC AAATAAATTG GAGCTTAACA TCCATTTCCA GCAATTTTTT AGAAACTACA 3540

GTGGACTATT CTAGATTCAA CATATTATAA AAACTAGAGT AAAAGAAAAG GATTGGATCT 3600

TGTGTAATGC AGGATCCAAT CCTTTCAATC ATTTTGTCCA ACTTTTGGAG GTTCCTACAA 3660

TGTAGTCGTC ATTAATAAAG ACAGATGGGA ATGACAGTGT TCCTATTTAT TTTGATAGAG 3720

ATCGATGAAT TCTTTAGATA GCAACTGAAT AATCTCTGTT GAAGCCATTT GGTCTTCTGC 3780

ATGCATAAAT AGCAAGGAGA ATCCTATTTT TTCTCCAGTA GCTTCTTTTT GTATGAGATT 3840

AGAGTGAATC TTGTGCGCTT CTACTAAGGA GTCTTCCGCT TCTTCAACTT TAATTTTCGC 3900

TTCTTTTAAA TTTCCTGCCT TAGCTAGTTG GATGGCTTCA ATAAAGGATG ATTTGGCTGC 3960

TCCACTATTG GCAATGAGCT GAAAACAGAT ATATTCCATT TCTTCTGTCA TCTTATTTCT 4020

CCTATCCATG CAAGTGCTTG TTCCAGAACT TTTGCTCCAT TCATCATTCC GTAATCCCGC 4080

ATATCAATGG TATCTACAGG GATATTTCCT GCAATTTCTT TCACAGCAAG TAACTCATAA 4140

CGAATTTGTG GCCCAATTAG AATGACATCT GCTTCATGGA TATTCTTTTT AGCTTCTGTC 4200

ATTGATTTTG CTTGGATAGA GATTTCAATC CCACGTTCAG TCGCACTTTG TTGCATTTTT 4260

TTAACAAGCA TACTTGTCGA CATTCCCGCA TTACATACTA ATAAAATTTG TTTCATAATC 4320

TTAACCTTCC ATTTCTTGTT CAACAACTTT GTCATTAACT TTGATAAATG GAATGTATAG 4380

AAGAACTCCA AGTGCAAAGA TGATGAATTG AACTAGAACT GCTCTCACGT CCCCTGCTGT 4440

TGCTAACCAT GCATTTAAGA ATACTGGTGT AGTCCAAGGA ACTTGTATAA ATGCAGGACT 4500

CATGAATTCT GTAACTGTTG CTAAGTAGCT GATTAAAATA CCAAGGACTG GAACTGTGAT 4560

AAATGGAATA GCTAATGAAA TGTTATAAAC GATTGGGTAA CCGAATAATA CTGGTTCATT 4620

GATATTGAAG ATACCAGGTC CAAAAGATAA TTTAGCCACG TTTTTAGAGA CAGCATTGCG 4680

ACTCACTAAG AATGTTGCTA TTAATAAACA TAATGTAGAT CCACTACCAC CCATTAAAGC 4740

GAATGTTTGT ATTTGTGATA GGTTGATGAT GTGTGGAATG GCTTGTCCAT TATTTGCTGC 4800

AGTGATGTTT TCAGTAATGT TAATTAATAG TAATGGTTCT AGGATGGCAC TGTAAATAAC 4860

TGCTTGGTGA ATACCAAATA GCCATAACAT ATTTCCTAAA GAGTAAATAA TAATGACCCC 4920

GATTAAGCTT GTACCAATAT GACGAATTGG TTCTTGAATA AAGATTGTAA TGATTGAGAT 4980

TAAGTTCATT CCAGTTATAT TGAATAATAA TGCTGAAACA ACCCCAAATA AGGAGATGAC 5040

GGTCATGACT GGAAGTAATA CGCTAAATGA TCTACTAACA GCTGGTGGAA TATTTTCACC 5100

AAGGTTCATT TGTAAAGCTT TAACGTTTGA TAATTCAATG AATAATTCTG TTGCAATAAT 5160

CGtACGATAA CCCCGGCGAA CATTGCGCCT GTACCTGTGT TGTTGAATGA AAGAACACCT 5220

GAAATGTTTA CCGCATCTTT TGCTCCGTCA GGAACTACAG AAACTGTATT TGGCATCATC 5280

ACAATTAAAG AAACTAATGA TAGCATTGAT GCTGCTAACG GGTTTTCGAA ATCTCTGTTT 5340

TTAGCTAAGA AATAACCAAC CATTACAGCA ATAATCATAC CTGAAATACT TAAAGTACCG 5400

TTTGCAATTG TTATTCCCCA ATATTGGAAT CTTGTTAATG TATCCCCTTG GAAAATCCAC 5460

TTAAATACCG TGTTGTTCAA AAGAACGATT AAACCTGCCA AAATATATAA TGGCATTACT 5520

GTTACGAATG CATCTCTTAG GGTTTTTAAA TGAATTTGGT TCCCTAGTTT ACCAGCAAAG 5580

GATGGCAAAA AAATTTTTTT GGGGGGGGGG GTTATTAAAC CCCCCTTTTT AAAAAAAA 5638

(2) INFORMATION FOR SEQ ID NO: 218:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4745 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 218:

CCGGAAGCTG TTGCCCTTGG AACTCCAAAT GAAGAAACAG CCTTTGTCTT GAACTATTTT 60

GGTGTGGAAG CACCACGTGT TATCACTTCT GCCAAAGCAG AGGGGGCAGA GCAAGTTATC 120

TTGACTGACC ACAATGAATT CCAACAATCT GTATCAGATA TCGCTGAAGT AGAAGTTTAC 180

GGTGTTGTAG ACCACCACCG TGTGGCTAAC TTTGAAACTG CAAGCCCACT TTACATGCGT 240

TTGGAGCCAG TTGGATCAGC GTCTTCAATC GTTTACCGTA TGTTCAAAGA ACATGGTGTA 300

GCTGTGCCTA AAGAGATTGC AGGTTTGATG CTTTCAGGTT TGATTTCAGA TACCCTTCTT 360

TTGAAATCAC CAACAACACA CCCAACAGAT AAAATCATTG CTCCTGAATT GGCTGAATTG 420

GCTGGTGTGA ACTTGGAAGA ATATGGTTTG GCAATGTTGA AAGCTGGTAC CAACTTGGCT 480

AGCAAATCTG CTGAAGAATT GATTGATATC GATGCTAAGA CTTTTGAACT CAACGGAAAT 540

AATGTCCGTG TTGCCCAAGT GAACACAGTT GACATCGCTG AAGTTTTGGA ACGCCAAGCA 600

GAAATTGAAG CTGCAATGCA AGCTGCCAAC GAATCAAACG GCTACTCTGA CTTTGTCTTG 660

ATGATTACAG ATATCGTCAA CTCAAACTCA GAAATCTTGG CTCTTGGTGC CAATATGGAC 720

AAGGTCGAAG CGGCTTTCAA CTTCAAACTT GAAAACAATC ATGCCTTCCT TGCTGGTGCC 780

GTTTCACGTA AGAAACAAGT GGTACCTCAA TTGACTGAAA GCTTTAATGC GTAAGATTTT 840

GGGTGTCAGC TCAAAATCGG AAAGTCTAGT TTGCCTTATA TCGCAAGGAG TTTCGGCTCC 900

TTTTTTCTAG GAGTGAAGTA TGTTAGAAAA TGGCGATTTG ATTTTTGTGA GAGATGGGTC 960

AGACATGGGA CAGGCCATCC AGACTTCCAC AGGTAACTAT AGCCATGTTG CCATTTATTT 1020

GGATGGGATG ATTTATCATG CTAGTGGACA GGCTGGTGTT GTCTGTCAAG AACCGGCAGA 1080

CTTCTTTGAG TCCAATCATT TATACGACCT CTATGTTTAC CCAGAAATGG ATATCCAGTC 1140

GGTGAAGGAA AGAGCTTGCA AACATCTTGG AGCACCCTAC AATGCTTCTT TCTATCCAGA 1200

TGCAGCTGGT TTTTACTGCT CCCAGTATAT AGCAGAAATC CTACCTATTT TTGAAACTAT 1260

TCCTATGAAA TTTGGAGwTG GGGAGCAGGA GATTAGTGAT TTTTGGAGGG AGTATTACAT 1320

AGAACTAGGT CTGCCTGTTC CTCTGAACCA AGCTGGTACC AATCCTAGTC AGTTGGCAGC 1380

ATCGCCTCTG TTACAATGTA AAGAAAGGAA TCTTCATGAT TCAGATTTTT AATCCATCTC 1440

GTTTGACGAG ACAGCCATTT TTGGAGAATT GATCCGCTAT CTGGATCAGT ATGAGGATGT 1500

GATTCTACGG GAAATTAAGG CTCAATTTCC AGATGTTGCA GTTGATAAAC TCATGGAAGA 1560

GTATATAAAG GCAGGCTTGA TTCTACGTGA AAATAAGCGC TATTACCTCA ATTTTCCTAC 1620

GCTTGAATCA CTTGATAGTC TTGAACTGGA TCAAGAGATT TTTGTCAGAG AAGCTAGTCC 1680

GGTCTATCAA GCCTTGTTGG AGCAGAGTTT TGAGACGGAA TTGCGCAATC AAATCAATGC 1740

AGCTATTTTA GTTGAAAAGA CGGACTTTGC GCGCATTAAA ATGACCCTGT CCAATTATTT 1800

TTACAAGGTC AAACAGCAGT ATCCTTTGAC AGAAAAACAG CAGGAGCTCT ATGACATTTT 1860

AGGAGATGTT AATCCTGAGT ATGCCCTCAA GTATATGACG GCTTTTTTGT TGAAATTTCT 1920

CAAAAAAGAC CAGCTTATGC AGAAATGCCG TGATATCTTT GTGGACAGTT AGGTTGTCTT 1980

AGGCTATATT GTGCAAAATG AAGATGGAAA GTATGAGTTG GCTATCGATT TTGATAAGGA 2040

GAGGTTAACT TTCTACTTAG CGTGATTTCT TGTTTCTGAG TACATTGTTT GACTTTCCTT 2100

AGTATTCGGT ATAAACTATA TGTAACCGGT AACACATATC GGAATAAACT AAAGGAGACA 2160

ATCATATGTC ACTTGAAAAC AAATTGGAAC AAGCAACAGG CGCTGTCAAA GAAGGTTTTG 2220

GTAAAGTTAC TGGAGACAGC AAGACAGAAC TTGAAGGAGC TGTTGAAAAA ACAGTTGCTA 2280

AGGCAAAAGA CGTTGTAGAA GACGCAAAAG GTGCTGTAGA AGGTGCCGTT GAAGGTTTGA 2340

AAAACGTTTT TACTAAAGAA TAGGAAAAAA TCAAGGGTTT CATTTTCCCT TGATTTTTTC 2400

TATTCTTATA AATAATTTTC TGCGACGGCT GTATCTCCTG GGTAGGATTC TTTCTTGCCC 2460

TGGATGATTT GGTAACAATC GGCTCCCTTA CCCGCAATAA TAACTGCATC TAATTCGTGA 2520

TTTGTGATAG CCATTGCCGC CTTGATGGCT TCTTGGCGAT CCGCAATCTT TTCAACAGGA 2580

TGATTGATGT AGCTACTAAT TTCATCTGCA ATGGCCATTG GGTCTTCATA GTTAGGGTCA 2640

TCAGCAGTCA GAAAGACTTG AATCTCAGGG TGTTGATTGA GGAGGAGGCC AAAGTCCTTA 2700

CGACGACTTT CTCCCTTGTT TCCTGTTGAT CCCAGAACCA GAGCAATCTT TCCGGTTTGA 2760

TGAGTTTCAA CCACATTGAT GAGTTTTTTC AGACTATCCC CATTGTGGGC ATAGTCGATG 2820

AAGACCTTGG CTCCATTTTT CTGAGTGAGG ACTTCCATAC GACCAGGAAC GCGGGTTGCA 2880

GCGATGCCTT TTTTGATGTC CTCAAGACTT GCTCCGAGAC GGAGACAAGC AAGTCCAGCA 2940

GCAACTGCAT TTTCTTGGTT GAAGTTGCCA ATGAGTTGAA TATCATAATC TCCAGCGAGT 3000

TTACCCGTAG CTGAAAAGCT AAAGGCTTTG GAATTCTCGA TTTGGTTATC AAATTGGCTA 3060

CCATAGAAAT CATGGTCTTG ATCTTCAACC TGTTCTTTCA AGACTGAGAA GTGGTCCATG 3120

TCACTGTTAA TGATGACTGC TCGGCTCTTT TCCATCAAGA GACGCTTGTG GTAGAAATAG 3180

TCTTCAAAGC TAGGGTGTTC AATCGGGCCG ATATGGTCTG GGCTGATATT TAGGAAAACT 3240

CCCACATCAA AGGTTAGACC ATAGACACGT TTGACCAGAT AGGCTTGACT GGAGACTTCC 3300

ATGATGAGGT GGGTACGGTC ATTTTGCACA GCCTGATTCA TCATGTCAAA GAGGTCAATA 3360

CTCTCAGGGG TTGTCAACGC TGACTTAAAG AAAGTCTCGC CATCAAGAGT TGTGTTCATG 3420

GTCGACAACA TAGCAGGTCT ATGCCCTTGA GATAAGATGT TATAGGCGAA ATAGGCTGCT 3480

GTTGTCTTAC CCTTAGTACC AGTAAAGGCA AGGAGTTTGA GTTTTTCCTG TGGATTACCA 3540

TAGAACTCCA TGGCAATCAA ACTCATGGCT TTCTTTATAT CGTTCACAAT GATGACAGGG 3600

ATACCGACTT CGTAGTCCTT TTCAGCTACA TACCAAGCTA ATCCTTGTGT TATAGCAGAA 3660

AGAAGGTATT CTTTTTTAAA GGCAGCGCCT TTTGCGAAAA AAAGAGTGTC TTCTGTTACT 3720

TTTCGGCTGT CGTAGCTGAT GCTATCAAAA ATAACTTTGC TGTAGTTGTA GTGGTAATGA 3780

CCTTGGTCAA TAATTTCGCG AAAAAGGCCA TCTTTCTTTA AAATATCTAA TACGGTTTCA 3840

ATCTTAATCA TACTTTCTAT TGTAAACCGA AAGTCGTAAA TTTACAAGTA ACAAGGAAAA 3900

GTTTATAATG GAAGATAAGG AGTTTTTCCT AGTTATCAAA ATTGAATGAG GAATCTATGT 3960

CGCACGAAAA CAATCACCAG CAGGCCCAGA TGTTACGGGG GACTGCTTGG CTAACGGCTA 4020

GTAACTTTAT CAGTCGCCTA CTCGGGGCTG TTTACATTAT CCCTTGGTAC ATCTGGATGG 4080

GGGCTTATGC AGCTAAGGCA AATGGTCTCT TTACCATGGG TTACAATATC TATGCTTGGT 4140

TCTTGTTGGT TTCAACAGCG GGGATTCCAG TTGCGGTGGC CAAGCAAGTT GCCAAGTATA 4200

ATACCATGCG AGAAGAAGAG CATAGCTTTG CCCTGATTCG GAGCTTCTTA GGCTTTATGA 4260

CAGGACTAGG CCTGGTTTTT GCTTTAGTCT TGTATGTCTT TGCTCCTTGG CTAGCAGACT 4320

TGTCTGGCGT GGGCAAAGAC TTGATCCCAA TCATGCAAAG CTTGGCTTGG GGAGTCTTGA 4380

TTTTCCCGTC TATGAGTGTT ATCCGAGGAT TTTTCCAAGG GATGAATAAC CTCAAACCCT 4440

ATGCCATGAG CCAAATTGCT GAGCAGGTCA TTCGTGTTAT CTGGATGCTC CTAGCAACCT 4500

TTATCATTAT GAAGCTCGGT TCAGGAGATT ATCTAGCAGC CGTTACCCAA TCAACCTTTG 4560

CTGCCTTTGT CGGTATGGTA GCCAGTTTTG CAGTCTTGAT TTATTTCCTT GCCCAAGAAG 4620

GTTCACTCAA AAGAATCTTT GAAACAGGAG ATAAGATTAA CAGTAAGCGT CTCTTGGTTG 4680

ATACCATTAA GGAAGCCATT CCTTTTATCC TGACAGGGTC TGCCATCCAG CTCTTCCAGA 4740

TTTTG 4745

(2) INFORMATION FOR SEQ ID NO: 219:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1900 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

CCTGATTGAC CTTATAATAA GGAACAAAAC ACAATGCACT ACCTTTTCAA CAAAAGAGTT 60

GCTGCTTGAT TAAAACCATC ACACCAGTTA TACCATTTTG CTTCATACCC ATCTTGAGCT 120

AGGATACGAT CTTCTAAATC AAAAACAGAG TAAATCTTTC TTTCCTCGCA AGCTTGCGCA 180

TAGAGATGAT ATAGTTCATC ACCACCATCT CTATCCCACT CAGCAGAAAT CGTATCCCGA 240

CCTGCCAATA AAGCCTGATA AGCCCTGTGA TGCCCATCTG TAATCAGCAA ACAATCTCCA 300

AAGGCAAGAA TACTGATTGG ATCGACTTGG ATTGTTTCTG CCGACTGGTA AAGCATCTGA 360

ATATCTTGCA ACTTCTTTTC TGATAAATAT AGTTGAGTCA GATGAAGATC TGCTATATTG 420

ACTTTCATTT CTTTCTCCTC AAGGGAATTC GATACTCACT TCTGTTTGCC TTTAAATCGC 480

CATTGGAAGC GGAgCTTGTC ATAAAAGGGA AACTCGATAA ACAGGACTCC CAAGCCCACA 540

CAGAGACTGG CAAGGACGTC TGATGGGTAA TGAACTCCCA GATAGACTCT TGATACCAGC 600

ACACTGACTA GGTAGAGGCC AAGGACGATT TGTACGATTT TTCTCCAGAC CTGATCTTTA 660

ATCCGCTGAC TAAGAATAAC AATCAAAGTC CCTACCATCA GCGTTACAGC TAGAGAATGC 720

CCACTTGGGA AGGAAAATCC CTTCTCCTCC ACCAGATGTA AAATAGCTGG TCGTGGGCGC 780

TGGTAGATAT TTTTAAAGGT CACGATTAAA AGACCTGCCA AAGCCAGATT TCCCAGCATG 840

AAGAAACTTT CTATCTTCCA TCGCTTACGA TAAAAGACAA AAGCTGTAAT GACAACCCAA 900

GTGATAATCA CTGGGATATC AATCAGACGT GTGAGGGCTC GAAAAAGAAT AGTCAAATAA 960

TCTGGTAAGT CTCCTCGAAT GGCAGTCTGA ATCGATTGGT CAAAATTGAC CAACATTTCA 1020

GGGTAAAATT TGACCATGTA GCCAAGAATA ACGAAAAGTA AAAGGGCAAA ACTGCCCTTC 1080

ATTAAAAATG TTTGTTTATC TCTCATAATG TTTTAAGGTT GGTTTCAAGA GAACATACAA 1140

CAACCAGAAT GAAACGGAAA AGATAACACC TTCAATCAAG TTAAAAGGTA ATACCATGGT 1200

CATTAGGTAG TTGGAAAGTC CCAAAATTTT TCCAATATCA AAGTTAGCAA ACTTAGCGTA 1260

CAAAGGAACA GCATAAACAT AGTTGAGAAC CAACATGGCC AAGGTTAAAC CAATAGTTCC 1320

AGCTAGAGAG CCTAGTAGGA AACGAAGGGT TGTCCGTTCC TTTTTCCAAA TCAAAGCAAA 1380

TACGATGACA AAAACTCCCA AAGCTACGAT ATTCATCGGC AAACCAATGT AAGTATTCAC 1440

TCCTTGGCTG TTAAGAAGCA ATTTCAAGAG TGAGCGAAGC AAGAGCACTC CTAGAGmCsC 1500

AGGCAAATCC ATGACCACCA GACCCACAAG GACTGGCAAG ATACTAAATT CGATCTTGAG 1560

GAAAGATGCC GCTGGTAAAA GCGGAAAGTC AAAGTACATC AGCACAAATG AGATGGCTGA 1620

TAGAATTGCA ATGGTCGAAA GTCGACGTGT GTTTGTCATA ACAGGTTCCT CCAATTTTCT 1680

ATAAAATCAG AAGAAGTTGG AAAGGATTCC TCTATCTATT CTCACTTTTT ATATCCCAAA 1740

AGTTCCCTCT TACTCTATTA AAGAAAAACA AAGCAAGTGG TTACAATCCG GCTATAAATC 1800

TATCAAAACA GACAAGGCTA TTCTTTCGTC TTCTCCCATC CAGACTATAC TGTCGGTTGT 1860

GGAATCTCAC CACATCACGT TGCGCTCACG GACTTCTTTA 1900

(2) INFORMATION FOR SEQ ID NO: 220:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4692 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

1230

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

GGTTTTCCAG CAGGAGCTTC TCCTTTATCA GAATGACCAT CCCATCTGCT CACGATAGAT 60

GAATAATGAT ATTTTTTACC ATGATAGTAA TTTGAAAAAG CCTAACCACC TCCTGAACCT 120

TCTCCATATG TCCATACTCC TCCATCTGGA TATTATACAG CAGCTGATGC AGCTCCCAAT 180

AATGTAAAAC TTGAAATAAG AGCTAGAGCA AGTAATCTAT GTTTTTTCGT TTTCATTTTA 240

TTTTTTCTTT CAAAAAAAGC ACACCTTGAG CAACAATGCA ACAAAATAAA TCCTCCTCTC 300

TCTTTTATTG AAACCGCTTT CTTATGTGAT AAGAATAACT TTTTTATtAT TTGTTGTCAA 360

GGAAAAAATC GAATTTTTTA GATATTTTAC TATATTACCT CTGTGAATAA TATTATATAG 420

TAGTTTTATT TCAAAATAAT ATGCAACCAG TACTAACCAA ATATAAAATA GATGCCATTA 480

ACGAATTTTA TTCAAGTTTT TCCCATTCAT ACTATACAAG TAAAAGAGAT GGTGTTAACT 540

AAAAAGCAAT TCAAACTATT GTAAAATTCC TAGCAAAAAG AGAGCCGAAA CTCTCTTTTT 600

TATCTTCTTT TACTTTTTTT GACTGGCATG AGTGTGATGT CTCTAACACT AAAGTAAGCT 660

AGGATCAACA TGGCTATTGC TAGGAATATT TCTGTTGGTA ATTGAAAAAT TTTCAGAAAA 720

GATAGAACCA ATAAAATCAA GAGTGCCACT AAAATACATA CCATAGCGAC GATATTGACA 780

GTCCCTTTAA TGCTTTCTGG TGTCGCAAAT ACATAGAGTA GGAGCAGTAA AATTCCTAGG 840

ACTAAATAGA CCATCTTTCT CTCTTTCTAG CTCTTATTCA GCTGATTTTT TCTTCTTGTT 900

AGCTTTCTCA CGCTCTGCTT TGTTAAGGAT TTGTTTACGC AAACGGATAG ACTCAGGCGT 960

TACTTCCATG TACTCATCGT CGTTCAAGAA CTCAAGAGAC TCTTCAAGTG TCAAGATACG 1020

AGGCGTCTTG ATAACAGCTG TTTGGTCCTT AGTAGCTGAA CGAACGTTGG TCATTTGTTT 1080

TGCCTTCGTG ATGTTAACTG TCAAGTCATT TTCACGAGAG TTTTCACCGA TGATCATTCC 1140

TTCATAAACC TCAGTACCTG GGTTGACAAA GATCGTACCA CGTTCTTCGA TAGACATGAT 1200

TGAGTAAGTT GTAGCCTTAC CAGCATCGAT AGAAACAAGG GCACCACGGT GACGTCCACC 1260

AATTTCCCCT GGAATCAATG GCAAGTATTG GTCGAAGGTA TGGTTCATGA TACCGTAACC 1320

ACGAGTCATT GATAAGAACT CAGTTGAGTA TCCAATCAAA CCACGCGCTG GAACAAGGAA 1380

GACCAAACGA GTTTGACCAT TACCAGTTGA AATCATATCC AACATTTCAC CTTTACGTTC 1440

AGAAAGGCTT TGGATAACAG ACCCTTGGTA TTCTTCTGGA GTGTCGATTT GTACACGTTC 1500

AAATGGTTCA CATTTAATAC CGTCGATTTC TTTTACGATA ACTTCTGGAC GAGATACTTG 1560

AAGTTCATAG CCCTCACGAC GCATTGTTTC GATAAGGATT GACAAGTGCA ATTCTCCACG 1620

TCCTGAAACA GTCCATTTAT CTGGTGAATC AGTTGGGTCA ACACGAAGGG AAACGTCTGT 1680

TTGCAATTCT GCCTGCAAGC GTTCTTCCAC CTTACGAGAA GTTACCCATT TACCTTCTTT 1740

ACCAGCAAAT GGTGAGTTGT TGACCAAGAA AGTCATTTGA AGAGTTGGCT CATCGATGTG 1800

TAGGATTGGA AGAGCTTCTA CTGCATCTGT CGGAGTGATG GTTTCACCGA CAAAGATGTC 1860

TTCCATACCT GAAACGGCAA TCAAGTCACC CGCTTTGGCT TCTTGGATTT CACGACGTTC 1920

CAAACCAAAG AAACCGAAGA GTTTTGTAAC ACGGAAGTTT TTAGTTGTAC CGTCAAGTTT 1980

AGAAAGGGTA ACTTGGTCCC CAACCTTAAC TGTACCACGG AAGACACGAC CGATACCGAT 2040

ACGTCCAACG AAGTCATTGT AGTCCAAAAG TGACACTTGG AACTGCAAAG GCTCATCTGA 2100

GTTATCTACT GGAGCTGGGA TATGGTCGAT AATCGTGTCA AAGATTGGTG CCATAGTCGC 2160

TTCTTGGTCA GCTGGATCAT CTGACAATGA AGAAGTTCCG TTGATCGCTG AAGCATAAAC 2220

CACTGGGAAA TCAAGCTGGT CGTCATCTGC ACCAAGCTCG ATGAAAAGTT CCAAGACTTC 2280

ATCCACTACT TCTGCTGGAC GAGCTGATGG CTTATCGATT TTGTTAACAA CCACGATTGG 2340

GACAAGGTCT TGTTCCAAGG CTTTTTTCAA TACGAAACGA GTTTGTGGCA TGGTTCCTTC 2400

ATAGGCATCT ACGACCAAGA CAACACCGTC AACCATTTTC ATGATACGCT CAACTTCTCC 2460

ACCAAAGTCC GCGTGTCCTG GTGTGTCCAT AATGTTGATA CGAGTTCCGT TGTAAGCAAC 2520

GGCAGTATTT TTAGCAAGGA TGGTAATTCC ACGCTCTTTT TCGATATCGT TTGAGTCCAT 2580

AGCACGCTCT GCCAATTCAG TCCGTGCATC AAGCGTTTCT GATTGTTTCA ATAATTCGTC 2640

AACCAGGGTT GTTTTACCGT GGTCAACGTG GGCGATAATC GCAATGTTAC GGATATCTTC 2700

TCTTAATTTT GTCATGATTT CCTCTATAAT ATTCAAAATT TATTTTCTAA CTGAACGATT 2760

ATACCATAAT TTCAAATAAA TAACATAACT CAAGCAAGTG TAAATGTTTT CACTCTGCTT 2820

TTCTTTTCAC GTCAAGCCTT TTCAAAGCGA GCGACTTATG ATAAGATAGG CACAGTATGC 2880

GTTTAGATAA TTTATTAGCT CAAGAAAAAA TCAGCCGAAA GGCCATGAAG CAAGCACTCC 2940

TCAGAGGGGA AATTCTAGTC GATGGTTGCC CAGCCCGCTC CCTAGCTCAA AATATCGATA 3000

CAGGACTACA AGAACTCCTT TTTCAGGATC GAATCATTCA AGGCTATGAA CACACCTATC 3060

TTATGCTTCA TAAACCTGCT GGTGCCGTTA CAGCCAACAA AGACAAGGAA CTTCCGACCG 3120

TCATGGACCT GCTTCCATCT AACATCCAGT CTGACAAGCT CTATGCCGTT GGCCGACTGG 3180

ACCGAGATAC AACGGGACTC CTCCTCTTGA CCGATAACGG TCCCTTGGGC TTTCAGCTCC 3240

TCCATCCCCA ATATCATGTC GATAAGACTT ACCAAGTTGA GGTTAATGGA CTTCTAACAC 3300

CTGACCATAT CCAAACCTTT CAAAAAGGAA TTGTCTTTTT AGATGACACT GTCTGTAAAC 3360

CCGCAAAACT AGAGATTCTA TCTGCAAGTC sCTCCCTCAG TCAAGCCTCT ATCACCATTT 3420

CAGAAGGAAA ATTTCATCAA ATCAAGAAAA TGTTCCTCTC GGTTGGTGTT AAGGTGACTA 3480

GCCTCAAAAG AATCCAATTT GGGGACTTCA CATTGAACCC AGATTTAGCA GAAGGTAACT 3540

ACCGCCCTTT GAACCAAAAA GAGTTACAAA TCATTAAAAA CTATTTAGAG ATGAGTCGAT 3600

AAAACAAAAA AAGCTTTAAA ACTAAAGCTT TTTTCTTTTA TTTACCGAAA AATTAAGGCG 3660

ATTGCTACAA TCCAGTTAAC TACAGAAATC ACAATTCCTA AGATATTAAG AATCTTTTCT 3720

ATTTTATAGT CTAATTGTGA CTCTTTTTGG TATGAAATAG CCAAGACCAA TCCTATGATA 3780

CCCAAAATCA GGCCTACAAT TGGAAATAAC AAACCAAGAA TAATCGACAA GATACCCACA 3840

AAAAGTGGAT TTTTCTTCTT TTCTTTTATG TTCTAAGAAC TCCTTAAATT TTATACAAAT 3900

TAATTATACT ATAAAACAAT AGCTTCATCC TATCATTCGA CTAATTTGGA AATAAGGTTA 3960

GCTAGTCTTC ACTTTCCCTT TCCAAGAATC CAAGCCATAA GAAAGGATAT AAATCTCAGA 4020

AAAACCTTGT TTTTTCAAGT AAAGAGCTGC ATTTGTAACT CGTTGCGCAC GTTGGTTTTC 4080

GTAGAGAAGG ACAGGTTTAT CTTTACGAAG GGCTGCAAGA CTAGTTTTCA ACTGACTTGA 4140

AGGAATATTG CGTGCACCAA GGATATGTTT TCTGTGGAAT TCTGCTGGGT CGCGCAAATC 4200

AATCAATTGA CCCGTACGAA TCAAGGCTTC AAACTCCTCA TTGTCCACAA TTTTAGCCGC 4260

ACGGCGAATA CGAAGATAGT TAAAGCCCAT CCACGCCAAC ATTGCTAGTA TAAGTGCCCA 4320

CAAAATCCAA GTAACCATTA GTTCTTTTCT CCATTTTTCT CAATATAATC CAATTCTACC 4380

TTGTGCTCTC TGCGAAGAAC TGCTTCTGCC TCTAGATAGT CTAATTTATC CATCAACCCT 4440

GCATCGTAAA TCCGAGATAG TTCCAACTTC ATCAGTTCAA TATCATATAA GCGTTTTCCC 4500

ATGTAAACAA TAATACCAAA TCGTTTGAGG AATTGCTGCA CATCATAGAA TGTTTTCATA 4560

AGACTCATTC TAGCAAAATT TTGTGTTTTT TTCAAGAAGA GACTCACACA ATGCTCCTTA 4620

TTTTCCTATC TTCTTTAGCG ATTCTAAGGC AAGTATGGTA CAATAAAAAC ATGGGGATTC 4680

AACAATTACA TT 4692

(2) INFORMATION FOR SEQ ID NO: 221:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 706 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

GCTAAAAAGC TGATAATCTT CGACTCCTGT ATATGATGTG TCTTTTCATG TAAGACACGC 60

GCCGCCAGAA TCATGGCAAG AGCTGCAAGA CTGGCAAGTA AGAAGCCGAT AAGATAGGCA 120

AAAAGATAAG TGAATTTGAC AAAGAAAGTC AAAAGAACTA GGAAACCAAA GCCTCCTCCA 180

AAAACTACCA AAGTCTTTCG TAAATCCCAG ATTTTATCCA ACTGCTTGAC GAGGGAAGTC 240

GTCTGACGAA CGCCTACAAT AGTTGCTAAC ATACTTCCTA AAAAGAATGG ATAGACATGA 300

GTTAAACTGG AGAAATAAAC AGAGGAATAA GAGGTCACTA GAAAACTACC AATAAACATG 360

GAGAAGAAAC TGATCAAGAA GGCAACAGCA GATAAGAGAA AGACCATCCC CTTCAACTGA 420

CCATTTGATT TAGCTTGTTT GGATAAGAAC CAAACTGCCA ATCCCCAAAG AATATAGTAG 480

TGAACCTCAA CTGCCAAACT CCAATTATGA ACAAACAAAT GAGGAATGAA CTGAGATTCA 540

TAACTCCCAC CTGTTAGGAG TTCATAGAAG TTGGTCATAA AGCCTAAGAC GCCCGCAATC 600

TGGCCACCAA TTCCAGCAAC ATAGTCTTGG CGAACCAAGA AAGTAAAAGG CATGGTCACC 660

AAGACCATCA AAACCACAGG TGGCACAATC TCGATAAAAG CGTCTT 706

(2) INFORMATION FOR SEQ ID NO: 222:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3236 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

CAGCTGATGG GCAATATCAG TCATAGAAAT TTTTTCAATT AACTTTTGAG CAATTTTTTG 60

GTTGATGATA CGAGGGATTT GGTGATTTTT CTTTACCAGG GGAGTCTCAG CAACCATCAT 120

TTTTGAACAG TGATAGCACT TGAAACGGCG TTTTCTAAGG AGAATTCTAG AAGGCATACC 180

AGTTGTTTCG AGGTAAGGGA TCTTAGACGG TTTTTGAAAG TCATATTTCT TCATTAGACT 240

TCCACAATCA GGGCAAGATG GAGCCTCATA ATCCAGCTTA GCGATAATTT CTTTGTGGGT 300

ATCCATATTG ATGATATCTA GAATCTTGAT GTTTGGGTCT TTAATATCGA GCAGTTTTGT 360

GATAAAATGT AATTGTTCCA TATGATTCTT TCTAATGAGT TGTTTTGTCG CTTTTCATTA 420

TAGGTCATAT GGGACTTTTT TTCTACACAA AAATAAGCTC CATAATATCC ATAGGGGATT 480

TACCCACTAC AAATATTATA GAGCCCGAAA ATATGGGAAA ACTGATCCTT GTTTCTGCTT 540

TTGTCTATAG AAGAATAATA AAGATTATCT TCTTCAAATT CTCCGATATT CTCTAAAGTT 600

TTGTGCAAGT TGCACAGAAC TTGTTTATTT TTTTGGTCAT CTTGCCATAG AAATATAAAG 660

CGTTTTCATA TATAATATAA TTATCAAAAG ACAAAAGGAG TTCACCTCAT GGTAGAATTG 720

AATCTTAAAA ATATTTACAA AAAATATCCA AACAGCGAAC ACTATTCAGT TGAAGATTTC 780

AACTTGAACA TCAAAGATAA AGAATTTATC GTTTTCGTAG GACCTTCAGG ATGTGGTAAA 840

1234

TCAACTACAC TCCGTATGAT TGCTGGTCTT GAAGACATTA CAGAAGGTAC TGCATCTATC 900

GATGGCGTAG TTGTCAACGA CGTAGCTCCA AAAGACCGTG ATATCGCCAT GGTATTCCAA 960

AACTACGCTC TTTACCCACA CATGACTGTT TATGACAACA TGGCTTTCGG TTTGAAATTG 1020

CGTAAATACA GCAAAGAAGA CATTAACAAA CGTGTTCAAG AAGCAGCTGA AATACTTGGA 1080

TTGAAAGAAT TCTTGGAACG TAAACCAGCT GACCTTTCAG GTGGTCAACG TCAACGTGTT 1140

GCCATGGGGC GTGCGATTGT CCGTGATGCG AAAGTATTCT TGATGGACGA ACCTTTGTCA 1200

AACTTGGATG CCAAACTTCG TGTATCAATG CGTGCTGAAA TCGCTAAAAT TCACCGTCGT 1260

ATCGGAGCTA CAACTATCTA TGTAACTCAC GACCAAACAG AAGCGATGAC ACTTGCAGAC 1320

CGTATCGTTA TTATGTCAGC TACTAAGAAC CCTGCTGGTA CAGGTACTAT CGGACGTGTA 1380

GAACAAATCG GTACTCCTCA AGAAGTTTAC AAAAATCCAG TTAACAAATT CGTTGCAGGA 1440

TTCATCGGAA GCCCAGCTAT GAACTTCATC ACCGTGAAAT TGGTTGGTAG CGAAATTGTT 1500

TCTGACGGTT TCCGTTTGAA AGTGCCAGAA GGAGCATTGA AAGTTCTTCG TGAAAAAGGC 1560

TACGAAGGAA AAGAATTGAT CTTTGGTATC CGTCCAGAAG ACGTGAATGC AGAACCTGCT 1620

TTCCTTGAAA CATTCCCAGA CTGTGTTGTA AAAGCGACTA TCTCTGTATC AGAACTGCTT 1680

GGTTCAGAAT CTCACCTTTA CTGTCAAGTT GGTAAAGACG AGTTTGTTGC AAAAGTTGAT 1740

GCTCGTGACT ACTTGCAAAC AGGTGCAACA GTTGAGCTTG GATTTGACTT GAACAAAGCA 1800

CACTTCTTCG ATGTAGAAAC TGAAAAAACA ATCTACTAAA ATAAATAAAA TTCAAAGCAC 1860

TACAAGAAAA GATATCTCTT TATCAATTGT AGTGGAGAGA TATCAGTTAA TCTAGGGAGA 1920

GAAACAAAAT GCTTCTCTCC TTTTTGCTAG AGAAGTCATA TTATGCATCT ATATTGTGAT 1980

GCTCTTTAAT ACTCTTCGAA AATCTCTTCA AACCACGTCA ACGTCGCCTT GCCGTACGTA 2040

TGATTACTGA TTTCGTCAGT TTTATCTGCA ACCTCAAAGA TGTACTTTGA GCAGCTTACG 2100

GCTAGTTTCC TAGTTTGCTC TTTGATTTCC ATTGAGTATT ATTTGTGGGT ACCATCTACA 2160

AGTGAAGCTA TATGCGTAAA CTACGTGAGC AATTGAATTC GAACTAGAGA GGTAATAATA 2220

AATTTATGCT ATAGTTATGG TGACTTGTAT GCTTTTGATT CTAGTTTATC AAATAATAGA 2280

TTAGAATTGT CAGATAATAT CATTTTGTGT TATAATGAAG AAAAAACAGA GGTGTTCAAA 2340

TGTCAGAAGC AGGTCATAAG TTTTTAGCAA AATTGGGGAA AAAACGCTTA CGTCCAGGTG 2400

GAAAGCGTGC CACAGATTGG TTAATTGCAG AAGGAGGATT TTCAAAAGAA AAGAGAATAC 2460

TAGAGGTTGC GTGTAATAGG GGAACTACAG CAATTGAGTT GGCACAGCGT TTTGGTTGCA 2520

AGATAACTGC TGTTGATATG GATGCTCAAG CTTTAGAAGT GGCTAAAAAA TCTGCTGGAA 2580

CGGCAGGTGT TGCTCATTTA ATCAGTTTTG AAAGAGCAAA TGCAATGAAA CTTCCTTATC 2640

AAGATGCTAG TTTTGATATT GTTATAAATG AAGCTATGCT GACTATGCAA GCCGATCAAG 2700

CTAAGAAAAA ATGTGTAATG GAATATCTAA GGGTATTAAA ACCTGGAGGT CTTCTCTTGA 2760

CACATGATGT GCTTCTTAAG GAAGCTAAAG AGTCTATCAG ACAGGAATTA TCACAAGCAA 2820

TTCATGTAAA TGTAGGTCCT TTAACTCAAG ATGGTTGGGA ACAGGTGATG ATAGAATCAG 2880

GTTATTGTGA TGTGAAAGCA TTGACTGGTG AAATGACATT AATGAAATTA TCGGGTATGA 2940

TTTATGACGA AGGTTTGCTA GGAACTTTGA AAATTTGTGT AAATGCTTGT AAAAAGGAGA 3000

ATAGAAAGCA GTTTTTAACT ATGTATAAAA TGTTTGCTAA GAATAAACAG AAATTGGGCT 3060

TTATTGCGAT GGCTAGTTAT AAATCGTCAA AACGTTAGAT AATTATTGAA GTTAACTTTT 3120

CCTTTTTTCT TTCTTAAAAA ATATGCTATA ATAGAGAGTA AAAAACTTTG AAAGAAAGAA 3180

AAAGATGAAT TTAAAAGATT ACATTGCAAC AATTGAAAAT TATCCAAAGG GTACCG 3236

(2) INFORMATION FOR SEQ ID NO: 223:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2885 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

CCTGACTTTT CAAATTGGTT AGTTTGCCAC ACTTGGTTTA TATGGTCGTG GAAAGCATGG 60

CTATTACTTC TCAAAGGGCG ATTTCTCACC CCATGAAAAG TGTCTATTTT TGTTTAGGTT 120

TGTAAGTTAA TTCATTGTCA CATATTACTC TTTAACTGAT TGAGTGAGTA CCGCTTATAT 180

TTGATGCCAA ACGCCTTAAA AGTGTTACCC TCAAGTCCTT TTAGAATACG GCTATAATTC 240

CGCTCATTGT AAACTATCTT AAGCTCATCA CTATCTAGGT TGGTATTAAA AATGGTATTT 300

TCACGATTGT TTAGCACGTC AAAGAGTAAA TCCTGCTCCC AGTCACTCTT AGGCTTAATA 360

ACAGCATTTT TTGCTCCTAA ATCATCAATA ATTAAGTAAT CAACAGACTT CATGAGTTCA 420

GTAGCTTCAA ACTCTGTAAG TGTTGCACCT TTACCATAAT TCCACCCCTC TTTAATTTGT 480

TTGATCATTT CGGTTAGGCT TACAAAAAGC ACACTCTTAG GTTCTCCTTT TGTCTTATAC 540

CCCTCATTTA TACCTTTGGC AATAGCAACT GATAAAAGTG TTTTTCCAAT CCCTGTACCT 600

CCTGTGATAA GCGTATTTCC CCTCATGCCA TCAAGATATT TTTGTACCTG ACCTTTTGCA 660

AATTCTAAAA ATCGCTTTTC TTCTGATGTT ACAGCATTAA AATCATCAAA AGTTTTAGTT 720

TTAAACTCAT CTGCTACATA GCTCTTATTG CTCATCAACA CATTATAAGT TTGCATATAT 780

AGTTTAGCAT TCAAATTATC AGCAATCGCA TCTTCTTCAT CTTGCTTTTT CTGTTCTTCT 840

TGGCATTGTT CACAATAGGG TGGGATACAG CGAACTTCTT TTATTGCCTC TCCGTTCTCA 900

TTCCACCCCA CTACTACATG TCTTTCTCCT TTGATTTGTG TTAGCTGTAT TTCATGCTTA 960

GGACACAATT CGTCTAGTTT AAATGTCTCA ATATTTCCTA AACTAGATTG TAATGATTTC 1020

ATTTTCTGAC CTCCTAAAAT GGTTTTTCTT GTGTTGGTAT CCAATCTTCA TAGCTGGTAG 1080

GCTCTAGTTG ATTGGTTTGC TGTTTTTTAG CCTCACGCGC TGCCCTGCTA TTTCTAACAA 1140

GTTCCACCGT CAATAAATTG TCCTGTTTCC AACGGTTAAG GATTACCTTG ATGTATGCAA 1200

AGTTTGCTTT ACCCTGACTG ACAGCCTCTT TTAACGCCTC ATGGATAAGC TCTGGGCTAA 1260

AATCTTCTAG CATATACTGC AATTCTTGAA TCTGTAACGG TGACAATGCT TTACCTGTCT 1320

CAGCTCGCTT CATATTCAAC AAGTCGTCTA TTTCCACACT GGTTACTTTT TTATTTACAA 1380

AATCAGAAAT CAGTTGAAAA ATGTTTGGAC TTTGTAGCTG GATTTCAGCC ATTACCTCAT 1440

CAAATTCTGC TTGTGTCATG TTGTCTAAAT CTAGTGTCAT TGCATTGCCT CCTCAAACTT 1500

CTCTATAAGA CAACTTTTAT TTGCTTTCTG AGTTCCATTT TTAGAGTTAA AAAGAATATC 1560

TTTTAAGGTT ACAGTAGCCT CTAAATACTC CTTTTCAGCA TGCTCTATAT ACGCCTGTTG 1620

CTCTGCTTCG TTCTCAAAAA AGTGCTTAGC TTGGCGTTTA AAGAATGCTT TTCGCATAGC 1680

GTCCATTTCA AAAATACCAG GGGCGAAAAA CATTCCCGTA GTGCTTTTAG AGACCGCTTC 1740

GATTTTATGG CTTTCATTCA ATTCAGGAAG TTCAATCCAA AGTAAACGGG ACAACTCATC 1800

TTTGATGGAT TTTGTCTGAC TTTCCAATAA AGAAAGGATT CTTAGGCCAT TTTCTTCGCT 1860

AATTTCTCGC ATTTCTGCGC TAATTCTGTC TATACGTCTA GTTAAATTCT CATATGTTGT 1920

TTCTGTCATG TTTTTACCTC TGTTTCTTTG TTGGTGTGAT TTTTTAGCTT ATTTTTTTAC 1980

TTCTAAACAT CATTGTCTTA ATTTCCTGAT AACTCATTTT CAATTCAATC ATAGCTATTG 2040

CCATATCCTC AAATGCCTGG TACTGCTCCA ACTCCTCACT AGTCAAGCTA TCGATACCGT 2100

TATAGCCCCC ACGCTCTTCT CTTAACTGCT TAGCGTTCAT GTCTGTTACT GCCTTTAGTA 2160

GCAAGTTGTT CATGGTGCTA TGCGCGTGCT TTGGTGCATT AGGCCATGTT TCTATACTGT 2220

CATGCAAGGT TTTTCTTTTC GGTTTTTCTA GCGCCCTCTG CAGACGAATT TCAGAAAGTT 2280

CCTCACGCAT TTCAAAGAAT GCTTTGACTA GGTTTAGTTT GAATTGCCGT ACTGTTTCGG 2340

TATTCTTTAA ATAAGTGATC AGAAAAGTAG CCTGTTGCTC GTTCAGAATA TAGGATTTTT 2400

TAGGTTGTCC TCTAGTATCT AATTTATGGA TTTTAAATCC AAGTATTCCC AACTCTTCAA 2460

AGTCAGCCTT ATTTTCTCTT ATTAAGCGCG TGATAGTGTG GTGTTGTACT TCAGCACATT 2520

CAGCGATGAT CTCGCTTGTG GTGTACGGCT CTTTCTTACC GTCCATGTAA ACTAGTTCCA 2580

TTACGGTTCT ACCTCCTGTA TAAATCTGGT TAGCTTACTT TTTAATTGCC TCCTCTAGCC 2640

TCTTTTTTAG CCTCTAAAAC GGCTTTGGCT AGTGGTTAAT ATTATTTACC ACTTGTCTCT 2700

ATAAACGTGT TAGAGGCCTT TATAACGACT TGTATCGCTG TATCGATATC CTCCGTGGAA 2760

TAGTAGATTT ATTTTCTAAT ATCATTCAAG ACTTGTTTAA CCCATTTCTT GAAAGAAATA 2820

AAATTACATC TTCTTTATCC TTGGCATCTG CTTTGTCTGA GACAAATTAG AATGTCAATA 2880

CTTGG 2885

(2) INFORMATION FOR SEQ ID NO: 224:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3144 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

TATCAATCCT TTCCCATTAT AGGAGCAACA GAGTGGGAGT AGTCATCTAA GGACTAATTT 60

ATGTATTTTT ACGAGTCAGT ATCTTGGGAT ACTGGTTTTT ACTTTTCTAG ACTTTTTGAC 120

TACTTGTTAA AACTGGGATA ATTTTCGACT GTTTAACAGT TATTATGCAA AGTCTAAAAG 180

ATTAGAATTG TCAAAACAAT CCGTCTAGGC TTGATTTTAT CCTTTATTTA CTATAAAATG 240

AGAAGGAAAA ATGTCAAACT TTTATATTGC AAATAGGAGA AATCATGACA AAAACATTAA 300

AACGTCCTGA GGTTTTATCA CCTGCAGGGA CTTTAGAGAA GCTAAAGGTA GCTGTTCAGT 360

ATGGAGCAGA TGCTGTCTTT ATCGGTGGTC AGGCCTATGG TCTTCGTAGC CGTGCGGGAA 420

ACTTTACTTT CGAACAGATG GAAGAAGGCG TGCAGTTTGC GGCCAAGTAT GGTGCCAAGG 480

TCTATGTAGC GGCTAATATG GTTATGCACG AAGGAAATGA AGCTGGTGCT GGTGAGTGGT 540

TCCGTAAACT GCGTGATATC GGGATTGCAG CAGTTATCGT ATCTGACCCA GCCTTGATTA 600

TGATTGCAGT GACTGAAGCA CCAGGCCTTG AAATCCACCT TTCTACCCAA GCCAGTGCCA 660

CTAACTATGA AACCCTTGAG TTCTGGAAAG AGCTAGGCTT GACTCGTGTC GTTTTAGCGC 720

GTGAGGTTTC AATGGAAGAA TTAGCTGAGA TCCGCAAACG TACAGATGTT GAAATTGAAG 780

CCTTTGTCCA TGGAGCTATG TGTATTTCAT ACTCTGGACG TTGTACTCTT TCAAACCACA 840

TGAGTATGCG TGATGCCAAC CGTGGTGGAT GTTCTCAGTC ATGCCGTTGG AAATACGACC 900

TTTACGATAT GCCATTTGGG AAAGAACGTA AGAGTTTGCA GGGTGAGATT CCAGAAGAAT 960

TTTCAATGTC AGCCGTTGAy ATGTCTATGA TTGACCACAT TCcAGATATG ATTGAAAATG 1020

GTGTGGACAG TCTAAAAATC GAAGGACGTA TGrAGTCTAT TCACTAyGTA TCAACAGTAA 1080

CCAACTGCTA CAAGGCGGCT GTGGATGCCT ATCTTGAAAG TCCTGAAAAG TTTGAAGCTA 1140

TCAAACAAGA CTTGGTGGAC GAGATGTGGA AGGTTGCCCA ACGTGAACTG GCTACAGGAT 1200

TTTACTATGG TACACCATCT GAAAATGAGC AGTTGTTTGG TGCTCGTCGT AAAATCCCTG 1260

AGTACAAGTT TGTCGCTGAA GTGGTTTCTT ATGATGATGC GGCACAAACA GCAACTATTC 1320

GTCAACGAAA CGTCATTAAC GAAGGGGACC AAGTTGAGTT TTATGGTCCA GGTTTCCGTC 1380

ATTTTGAAAC CTATATTGAA GATTTGCATG ATGCTAAAGG CAATAAAATC GACCGCGCTC 1440

CAAATCCAAT GGAACTATTG ACTATTAAAG TCCCACAACC TGTTCAATCA GGAGACATGG 1500

TTCGAGCTCT TAAAGAGGGG CTTATCAATC TTTATAAGGA AGATGGAACC AGCGTCACAG 1560

TTCGTGCTTA ATGTAGTTGT TTAGTTTTAA AAAACTATGC AAAGCTCCAT ATACAACACT 1620

TAAACGAGAT TAAAGAATGG CGAAATCCCT TGATGCGCAA GAGATTAGCT GTCTTTTTTA 1680

TTTTTTAAGT GATAAAGTCG GAGTTTAGGC ATCAAAGCCT ATCAAATTAA ACAAAGAAGC 1740

GATGTCTTAG ATATTTTGAA AAAAATTAAT AAGCAGAAAA CTCTCTATTA TTTTGTTGTA 1800

GAGAGTTTTT TGTTAATAAA ATTTCACAAA ATGACATTTA TATATTGCAT TAAGTTAGAT 1860

ATATGATATA ATATTGTTAA AAAGAGGCGC AACTTTTTAA AATTAATGAG AATCAAAGAG 1920

AAAACCAATA ATATTAATGG AGGAATAAAA AATGTAAGTA AGCATTATGG TCATTCAATC 1980

ATTCTCAAAG ATATAAATTT TGCACTTAAC AAGGGTGAAA TTGTTGGTCT AGCAGGGAGA 2040

AATGGAGTTG GTAAGAGTAC GTTGATGAAA ATTCTTGTTC AGAATAATCA ACCGACTTCA 2100

GGTAATATTA TAAGCAGTGA TAATGTTGGG TATTTAATCG AAGAACCAAA ATTATTTTTA 2160

TCTAAAACAG GTTTAGAGAA TTTAAAATAT TTGTCAAATT TATATGGTGT TGACTACAAT 2220

CAAGAAAGAT TTAGATGTTT GATCCAAGAG TTAGATTTGA CTCAGTCTAT TAATAAAAAA 2280

GTAAAGACCT ATTCTTTGGG TACAAAACAA AAATTAGCTT TGCTTCTAAC TCTCGTTACG 2340

GAACCTGATA TATTGATTTT AGATGAACCG ACTAATGGTT TAGATATTGA ATCATCACAA 2400

ATAGTTTTAG CGGTTCTAAA AAAATTAGCT TTACATGAAA ATGTGGGAAT TTTAATATCG 2460

AGTCATAAAT TAGAAGACAT TGAAGAAATT TGTGAGAGAG TTCTTTTCTT GGAGAACGGG 2520

CTTTTGACAT TTCAAAAAGT AGGAAAAGAT AGTCATAATT TCTTGTTTGA GATAGCTTTT 2580

TCATCAGCTA CAGATAGAGA CATTTTCATT ACCAAACAAG AATTTTGGGA TATTGTTTAG 2640

GAAGAGGGAT TGAGAATTAC TATGTCTGGG AATATTCAAA ATAGTGAGCT TTTTAAATTT 2700

TTTAACGAAA ACTCTATTAA AGTAGTTGAT TTTGAAACTA AAAAAGAGAC GCTTAAAGAT 2760

ATTTACCTAA ATCGTTCAAA ATAAAGGAAG GTTATAATCA TGAAATTAAA TAAACAGAAG 2820

AATCGGATGA TTTACGTCTT GTCTAATTTT CTATATGCTA TCTCAGTTTC CATTATTTAT 2880

GCTTTGAATG GCATTGTGTT ACTAGTCATA GTAAGTAAAT TGGGTATTCC AGGTGATTTA 2940

GGATTAAATT TTATAGTAGC TATTGTAGTC AATACAATTT TGTTAGTCCT GTTTTATTTT 3000

CTATTATCTT ACATTTTCTA TTTATACAAA TTGAAAAGTG GCTTGGTATw TGGTATTTTA 3060

GTAGCTTTAC TACTCTTTAT CTCTAATATA TTAAATACGA TGATGATGAA TACTAGTAAT 3120

GATTTGTTTA TCAAAGCAAT TGAA 3144

(2) INFORMATION FOR SEQ ID NO: 225:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3766 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

TACGGTATTA TTTTTAAGGA GAAAGAATCA TGAAAATCAA AAAATGGCTT GGTCTAGCAG 60

CCCTTGCTAC AGTCGCAGGT TTGGCTCTTG CAGCTTGCGG AAACTCAGAA AAGAAAGCAG 120

ACAATGCAAC AACTATCAAA ATCGCAACTG TTAACCGTAG CGGTTCTGAA GAAAAACGTT 180

GGGACAAAAT CCAAGAATTG GTTAAAAAAG ACGGAATTAC CTTGGAATTT ACAGAGTTCA 240

CAGACTACTC ACAACCAAAC AAAGCAACTG CTGATGGCGA AGTAGATTTG AACGCTTTCC 300

AACACTATAA CTTCTTGAAC AACTGGAACA AAGAAAACGG AAAAGACCTT GTAGCGATTG 360

CAGATACTTA CATCTCTCCA ATCCGCCTTT ACTCAGGTTT GAATGGAAGT GCCAACAAGT 420

ACACTAAAGT AGAAGACATC CCAGCAAACG GAGAAATCGC TGTACCGAAT GACGCTACAA 480

ACGAAAGCCG TGCGCTTTAT TTGCTTCAAT CAGCTGGCTT GATTAAATTG GATGTTTCTG 540

GAACTGCTCT TGCAACAGTT GCCAACATCA AAGAAAATCC AAAGAACTTG AAAATCACTG 600

AATTGGACGC TAGCCAAACA GCTCGTTCAT TGTCATCAGT TGACGCTGCC GTTGTAAACA 660

ATACCTTCGT TACAGAAGCA AAATTGGACT ACAAGAAATC ACTTTTCAAA GAACAAGCTG 720

ATGAAAACTC AAAACAATGG TACAACATCA TTGTTGCAAA AAAAGATTGG GAAACATCAC 780

CTAAGGCTGA TGCTATCAAG AAAGTAATCG CAGCTTACCA CACAGATGAC GTGAAAAAAG 840

TTATCGAAGA ATCATCAGAT GGTTTGGATC AACCAGTTTG GTAATAAGAA ACAGGGAGGT 900

GGGAGAGAAA ATTCCACCTC TTGCTTTTGT ATAGAGTATA GATTGTAAAG AAGACTATTC 960

GTTCATAGAA AGGTAGAGAG AATATGGTTT TTCCTAGCGA ACAAGAACAG ATTGAAAAAT 1020

1240

TTGAAAAGGA TCATGTAGCC CAGCATTATT TTGAGGTTTT GCGTACCTTG ATTTCTAAGA 1080

AATCAGTCTT TGCCCAGCAG GTTGGACTCA AGGAAGTCGC AAATTATCTG GGTGAGATTT 1140

TCAAGCGTGT TGGAGCTGAA GTGGAGATTG ATGAGAGCTA TACAGCGCCC TTTGTCATGG 1200

CACATTTCAA GAGTTCGCGT CCAGATGCCA AGACCTTGAT TTTCTATAAC CACTATGACA 1260

CTGTGCCAGC GGATGGGGAT CAGGTCTGGA CAGAGGATCC kTTTACGCTT TCGGTCCGCA 1320

ATGGCTTCAT GTATGGGCGT GGGGTTGATG ACGACAAGGG TCATATCACA GCTCGCTTGA 1380

GTGCTTTGAG AAAATATATG CAGCACCATG ATGATTTACC TGTCAATATC AGCTTTATCA 1440

TGGAGGGAGC GGAGGAATCG GCTTCAACAG ACCTAGATAA GTATTTGGAA AAGCATGCAG 1500

ACAAACTCCG TGGGGCGGAT TTGTTGGTCT GGGAACAAGG GACCAAAAAT GCCTTGGAAC 1560

AGCTGGAAAT TTCTGGTGGC AATAAGGGGA TTGTGACCTT TGATGCCAAG GTAAAAAGCG 1620

CTGATGTGGA TATCCACTCG AGTTATGGTG GTGTTGTGGA ATCAGCTCCT TGGTATCTCC 1680

TCCAAGCCTT ACAGTCTCTT CGTGCTGCGG ATGGCCGTAT CTTGGTTGAA GGCTTGTACG 1740

AAGAAGTACA AGAGCCCAAT GAACGAGAAA TGGCCTTGCT AGAAACTTAT GGTCAACGAA 1800

ACCCAGAGGA AGTTAGTCGG ATTTATGGAT TGGAGTTGCC TCTCTTACAG GAGGAGCGGA 1860

TGGCCTTTCT AAAACGTTTC TTTTTCGATC CAGCGCTTAA TATCGAAGGA ATCCAGTCTG 1920

GTTATCAAGG TCAGGGTGTT AAGACTATTT TACCTGCAGA AGCCAGTGCC AAGCTAGAGG 1980

TTCGTCTGGT TCCGGGCCTA GAACCGCATG ATGTTCTGGA AAAAATTCGG AAACAGCTAG 2040

ACAAAAATGG CTTTGATAAG GTAGAATTAT ACTATACCTT GGGAGAGATG AGCTATCGAA 2100

GCGATATGAG CGCACCAGCC ATTCTCAATG TGATCGAGTT GGCCAAGAAA TTCTATCCAC 2160

AGGGCGTTTC AGTCTTGCCG ACGACAGCGG GGACAGGACC TATGCATACG GTCTTTGATG 2220

CCCTAGAGGT ACCAATGGTT GCATTCGGTC TAGGAAATGC CAATAGCCGA GACCACGGTG 2280

GAGATGAAAA TGTGCGAATC GCTGATTATT ACACCCATAT CGAATTAGTA GAGGAGCTGA 2340

TTAGAAGCTA TGAGTAGAGA TATTATCAAG TTAGATCAGA TCGATGTGAC TTTTCACCAA 2400

AAGAAGAGAA CCATCACAGC GGTTAAGGAT GTGACCATTC ACATCCAAGA AGGGGATATC 2460

TACGGAATCG TTGGATATTC TGGAGCAGGA AAATCAACCC TTGTACGGGT GATTAATCTC 2520

TTGCAAAAAC CATCTGCAGG GAAAATTACC ATTGACGACG ATGTGATTTT TGACGGCAAG 2580

GTGACCTTGA CGGCAGAGCA GTTGCGTCGT AAACGTCAAG ATATCGGAAT GATTTTCCAG 2640

CATTTTAACC TGATGAGCCA AAAGACAGCA GAGGAGAATG TAGCCTTTGC CCTTAAACAC 2700

TCTGAACTCA GCAAGGAAGA AAAGAAGGCT AAAGTAGCTA AGTTGTTGGA CTTGGTTGGT 2760

TTGGCAGATC GTGCTGAAAA CTACCCTTCA CAACTATCTG GAGGGCAAAA ACAGCGTGTG 2820

GCAATTGCGC GTGCCTTGGC CAATGATCCA AAAATCTTGA TTTCAGACGA GTCAACTTCT 2880

GCCCTTGATC CGAAGACAAC CAAGCAGATT TTGGCCTTGT TGCAAGATTT GAACCAAAAA 2940

TTAGGCTTGA CTGTTGTCTT GATTACGCAT GAAATGCAGA TTGTCAAAGA CATTGCCAAC 3000

CGTGTTGCAG TTATGCAGGA TGGGCATTTG ATTGAAGAGG GTAGTGTGCT TGAAATCTTC 3060

TCAAACCCTA AACAACCTTT GACTCAAGAC TTTATCTCAA CAGCTACAGG TATTGACGAA 3120

GCCATGGTCA AAATCGAGAA GCAAGAAATC GTGGAACACT TGTCTGAAAA CAGTCTCTTG 3180

GTGCAACTCA AGTACGCTGG AGCTTCAACA GACGAGCCAC TTTTGAATGA ATTGTACAAG 3240

CATTACCAAG TAATGGCTAA TATTCTCTAT GGGAATATCG AAATTCTCGA TGGTACTCCT 3300

GTTGGAGAAT TGGTGGTGGT TTTGTCAGGT GAAAAAGCAG CGTTGGCAGG TGCCCAAGAA 3360

GCCATTCGTC AAGCAGGTGT ACAACTAAAA GTATTGAAGG GAGTACAGTA AGATGGAATC 3420

ATTGATTCAA ACCTATTTAC CAAATGTCTA TAAGATGGGT TGGGCTGGTC AGGCAGGCTG 3480

GGGAACGGCT ATCTACTTAA CTCTTTATAT GACAGTTCTT TCCTTCATTA TCGGAGGCTT 3540

CTTGGGGCTA GTGGCAGGTC TCTTTCTCGT CTTGACAGCG CCAGGTGGTG TCTTGGAGAA 3600

TAAAGTCGTA TTCTGGATTT TAGACAAAAT TACCTCAATT TTTCGTGCGG TTCCCTTTAT 3660

CATCCTCTTG GCAATCTTGT CACCACTTTC TCACTTGATT GTTAAAACAA GTATCGGGCC 3720

AAATGCAGCC CTTGTCCCAC TTTCTTTTGC AGTCTTTGCC TTCTGG 3766

(2) INFORMATION FOR SEQ ID NO: 226:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2520 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

TGTTGCTGAG TTAATCGGTA CGTTCATGTT TGTATTCGTC GGGACAGGAG CTGTTGTTTT 60

TGGAAATGGT CTTGATGGCC TTGGTCACCT TGGAATCGCC TTTGCCTTTG GTTTGGCAAT 120

CGTGGTGGCA GCCTACTCAA TCGGAACTGT TTCAGGTGCT CACTTGAACC CGGCTGTTTC 180

GATTGCTATG TTTGTAAACA AACGTTTGTC ATCTTCAGAA CTTGTAAACT ACATCCTTGG 240

TCAGGTTGTT GGAGCTTTCA TCGCTTCTGG CGCTGTCTTC TTCCTCTTGG CTAACTCAGG 300

TATGTCAACT GCTAGTCTTG GTGAAAATGC CTTGGCAAAC GGTGTCACTG TCTTTGGTGG 360

TTTCTTGTTT GAAGTCATCG CAACTTTCTT GTTTGTATTG GTTATCATGA CTGTGACTTC 420

AGAAAGCAAG GGCAATGGCG CGATTGCTGG TTTGGTAATC GGTTTGTCAT TGATGGCGAT 480

GATTCTTGTC GGATTGAAGA TTACTGGACT TTCAGTAAAC CCAGCTCGTA GCTTGGCACC 540

AGCTGTCTTG GTAGGCGGCG CAsCCTTCAA CAAGTTTGGA TTTTCATCCT TGCACCAATC 600

GCTGGTGGAG TTCTTGCAGC CCTTGTTGCA AAAAATTTCC TTGGAACAGA AGAATAATTG 660

AAACTCAAAA AGCCTTGCTC CTCATCTTGA GGAACAGGGC TTTTTCGTAT GATACTCTTC 720

GAAAATCTCT TCAAACCACG TCAGCTTCAT CTTGCCGTAG TATGGTTACT GACTTCGTCA 780

GTTCTATCCA CAACCTCAAA ACAGTGTTTT GATCTGACTT CGTCAGTTCT ATCTGCAACC 840

TCAAAACAGT GTTTTAAGCT GACTTCGTCA GTTCTATCTG CAACCTCAAA ACAGTGTTTT 900

AAGCTGACTT CGTCAGTTCT ATCTGCAACC TCAAAACAGT GTTTTAAGCT GACTTCGTCA 960

GTTCTATCTG CAACCTCAAA ACAGTGTTTT AAGCTGACTT CGTCAGTTCT ATCCACAACC 1020

TCAAAACAGT GTTTTGATCT GACTTCGTCA GTTCTATCCA CAACCTCAAA ACAGTGTTTT 1080

GATCTGACTT CGTCAGTTCT ATCCACAACC TCAAAACAGT GCTTTGAGCA ACcTGCGGCT 1140

AACTTCCTAG TTTGCTCTTT GATTTTCATT GAGTATGACT TTAGCGGTTG TCAATTTTCT 1200

CTGGATAAAG GTCGTGTTGG AAGAGGCGTT GTTCTGCCAA GCCCTCATAC TTAGTTCCTT 1260

GCTTACCGTA GTTGTAGTAG GGGTCGATTG AAATGCCACC GCGCGGAGTG AATTTTCCCC 1320

AGACTTCTAA ATAGCGAGGG TCTAGCAAGT TGACCAAGTC TTTCCCGATG GTGTTGATAC 1380

AGTTTTCGTG GAAATCTCCG TGGTTTCGGT AGCTAAATAG ATATAGTTTG AGGGATTTTG 1440

ACTCGACACA GAGCTTGTCA GGAATGTAGG AAATATGAAT CGTCGCAAAG TCTGGCTGAG 1500

CAGTGATTTG TCCCAGCAGA GACATATCGA GGATATGGTG ACGAATGCCC TGTTCCTTAG 1560

CGATTTCTCT AGTAATTTGA ATTTCGAGGT GATGACGTTG GCCGTAGGCA AAGGTGACAG 1620

CTTCGACTGT TTCATAGTGT TGCATGACCC AGAAAAGGCA GGTTGTTGAA TCTTGACCAC 1680

CACTAAAGAC GACCAAGGCT AATTGACGTT TCATAGTACT CCTTCCAAAA TGGGAAATGT 1740

TCAGAGCACG CAAAAAGCTC CCATTAGGGA GCTAAAAAAT ACCAAATCGA GGTTTTTTTA 1800

GCGATGGCAT ATCCCAAACA TCGTAATATT CTACTTATAT AGTAAAATGA AATAAGAACA 1860

GGACAAATCG ATCAGGACAG TCAAATCGAT TTCTAACAAT GTTTTAGAAG TAGAGGTGTA 1920

CTATTCTAGT TTCAATCTAC TATAGTCTAG CATATTTTTT GAAAAATGGC AAAGGGCAAG 1980

AAAAAAGAGA CCAAAGAAAG TACTTGGTCT CTCGTTTGAT TAGCTCAATT CAGCAATGAT 2040

GGCCTTGATT TGTTCTGCTG TGTGAACACC TGCAACTTGT TTGACAACTT GGCCGTCTTT 2100

TTTGAAGAGA AGAGTTGGAA TAGACATGAT TCCAAAAGCA CGAGCTGTGT TTGGATTTTC 2160

ATCAACGTCC ATTTTAACGA TTTTCAAGAC ATCTTCTGAA AGTTCTTCAG ACAATTTGTC 2220

CAAGATTGGA CCTTGCATAC GACATGGACC ACACCAAGTT GCCCAGAAGT CTACTAAGAC 2280

CAAACCGTCT TTTGTTTCTT GTTCGAATGT TGCATCTGTA ATTGCTTTTG CCATTGTATT 2340

TCTCCTTTTT TTAGTTATAT TGGCTTAAAT CTTGTTTCAT GAGATAGAAG AAGATATCTC 2400

CATAAGTCCC ATGGTAGTCC AAATTATGAC CCTTGTAAGT TAATTTTTGG ACAGGGTAGT 2460

AkkCTGCGAC GCCGATAAGG CAAGCTTGTT GCGAACGTTC AAAGTCTTCA TAAGACTCGG 2520

(2) INFORMATION FOR SEQ ID NO: 227:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5278 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

ACTCAGTTAG ATTTTGTTTT CAAAAACAAC GAAGAAAAAG ACCATGTTGC TCTACTTGGA 60

AGAATTGGCT CCGAACGTGT TTATCGATAT ATTAATAAAA AATATTTAGA TTTACCGGAA 120

ACATTCGAAA ATTATAATGT TTTTGTACCA GAAGCTAATG GAAGTGGTGC CTTAGGTGAA 180

GTCTTATCAA CACCCCTAAT CGGGGAACCC CTAATCGGGC ATACAGATAC TTTTTTATCT 240

ATTGGTAATT TTAAAACAAA ATTTGAAGCC GATGCTTGTA TTAAATTTAT TAAAACTAAA 300

TTCGCTAGAG TATTATTAGG TGTTTTGAAA GTTACTCAGC ATAATTCACG CAAAACTTGG 360

TATTACGTCC CCCTCCAAGA CTTTACGGTC AATTCGGACA TTGATTGGAC ACAATCAGTG 420

ACTGATATTG ACCGCCAGCT TGATCAAAAA TATGACTTTT CCCCTGAAGA AATTGCCTTT 480

ATTGAGAATC ATGTAAGGGA GATGGATTAG AAAAGTATTT TTATTTGACA AATAGTGCTC 540

AATGATCTAA AATGACTATA TAGGATTAGG TCAGGAAGCA TACGATGCCC TGACCCTTTT 600

TGTACTTATG AGATGAGAAA GTCATTTGTT AGATAAATTG ACTCGTTAGC AAACGTTCAA 660

AAAAGGAAAA CTTATGCCAG TAGAAATTAA AACCACTAAA GAAATTCATC CTAAAATCTA 720

TGCCTACACC ACACCGACAG TAACCAGTAA TGAAGGCTGG ATTAAGATTG GGTATACAGA 780

ACGTGATGTC ACACAACGTA TCAAGGAGCA AACGCATACA GCTCATATAG CTACAGATGT 840

CTTATGGACT GGTGATGCAG CTTATACAGA AGAGCCTGAT AAGGGGAAAA CTTTCAAGGA 900

CCATGATTTC CACCATTTCC TTTCTTTCCA TGATGTAGAA CGTCGTCCCA AGACGGAATG 960

GTTCTATTTT AATGGAACTC CTGAAAAATC AAAAAATCTT TTTGATAAGT TTGTTCAGCA 1020

TGATTTGTCT GGTTATCAGC CTGGAAAAGG ACAGGACTAT ACTCTGCGAC AAGAGCAAGA 1080

1244

AGAAGCAGTT GCTAAGACAT TAGCTTATTT CCAAGAACAT GCTGGAGGCA AGTTTCTCTG 1140

GAATGCCAAG CCACGCTTTG GTAAAACCTT GTCTACCTAT GACCTAGCTC GACGGATGGA 1200

AGCTGTCAAT GTCCTAATTG TAACAAACCG CCCTGCCATT GCTAACTCAT GGTATGATGA 1260

TTTTGAAACA TTCATAGCAG GTCAAACGAC TTACAAGTTT GTTTCTGAAT CAGATAGCCT 1320

TAAGAGTCGT CCAATCTTGT CACGACAAGA ATTTCTTGGT ATTTTAGCTG ACGATGTAAG 1380

ACAACTTGCT TTTATCAGTC TCCAAGACTT GAAAGGATCT GTTTATTTAG GTGGAGAGCA 1440

CGATAAACTC AAATGGGTAA CTGATCTGCA TTGGGACTTG TTGGTTATTG ACGAGGCTCA 1500

TGAAGGAGTT GATACCTTCA AGACTGACCA AGCCTTTAAT AAGATTCGAC GAAATTTTAC 1560

TCTGCATTTG TCAGGTACAT CATTTAAAGC ATTGGCTAAA GGAGATTTTA CAGAGGAACA 1620

AATCTACAAC TGGTCTTATG CTGATGAGCA GGCTGCTAAG TATTCGTGGT CTCTTGAGCA 1680

AGAAGAGGAA AATCCTTATG AAAGCTTGCC TCAGTTGAAT CTCTTTACCT ATCAAATGTC 1740

TCAGATGATT GGCGAAAAGT TAGAAAAAGG CGCTCAGATC GATGGTGAAA ATATTGACTA 1800

TGTTTTTGAC TTAAGTGAAT TTTTCGCTAC AGATGATAAA GGGAAATTTA TTCATGAGCA 1860

TGATGTCAGA AATTGGTTAG ATACTCTATC AAGCAATGAA AAATATCCAT TTTCAACCAA 1920

AGAACTCCGT AATGAACTCA AGCATACTTT TTGGCTTTTA GAACGTGTCG CTTCGGCCAA 1980

AGCATTAAAA GCCCTACTAG AAGAACACCC AATCTATGAA AACTATGAGA TCGTTCTAGC 2040

TGCTGGTGAC GGACGTATGT CCGAAGAAGA CGATAAAGTC AAACTCAAAT CCTTGGACTT 2100

GGTTAGAAAA GCGATAGCAG AGAATGACAA AACCATTACC CTATCCGTTG GTCAGCTGAC 2160

GACAGGTGTC ACTATCCCTG AATGGACAGG TGTATTGATG TTATCAAATT TGAAATCACC 2220

AGCTCTTTAT ATGCAGGCCG CCTTCCGTGC TCAAAATCCT TACTCATGGA GCGATAACAA 2280

AGGAAATCAC TTTCGCAAAG AAAGAGCCTA TGTATTTGAC TTTGCGCCGG AAAGAACCTT 2340

GATTCTCTTT GATGAGTTTG CCAACAACTT ATTGCTTGTA ACTGCAGCTG GTAGAGGAAC 2400

TTCAGCTACA CGCGAAGAAA ATATTAGAGA ATTATTAAAC TTCTTTCCAA TTATTGCCCA 2460

AGACCGTGCT GGTAAGATGG TTGAAATTGA TGCAAAGGCA GTTCTAACCA CTCCTCGCCA 2520

GATAAAAGCT AGAGAAGTTC TTAAACGAGG TTTTATGTCC AATCTCTTAT TTGATAATAT 2580

TAGTGGTATT TTCCAAGCAA GTCAAACAGT TTTAGATATT TTAAATGAGC TGCCAGTTGA 2640

AAAGGAAGGG AAGGTACAAG ATAGTTCTGA TTTATTAGAT TTTTCAGATG TTACAGTCGA 2700

TGATGAGGGA AATGCAGTAG TAGACCATGA AATTGTAGTT AATCAGCAAA TGCGACTTTT 2760

TGGTGAAAAA GTTTATGGAC TTGGTGAATC TGTTGCTGAG TTAGTCACAA AAGATGAGGA 2820

ACGAACTCAA AAACAGCTGG TCAATGACTT GAGTAAGACC GTTTCTTCAG TGATTGTAGA 2880

GGAATTGAAA GCAGATTATT CTCTAAAAAC AAGGGAAACT GAGCAAATTA AGAAACAAAT 2940

TACAGCAACA CTTGAGAATG AAATTCGAAA AAATGATATC GAAAGAAAAA TTTCTGAAGC 3000

TCATATCAAG CAAGAGTTGC AACAGCAGCT CAAAGAAGCA AATGATAAAG CGCAAAAAGA 3060

TAAGATTCAA GAAGATTTGG AAAAACGTTT AGAAGAAAAT AAACTCATTC ATAAAGAAAA 3120

ACTAGAACAA ACACTCAAAA AAGAAGTGGA AAAAATGCCT GAGAAATTTA TCGAACAGGT 3180

TGAGATAAAA CGTGTGGAAC AGTTGAAACA ATCAGCTCAA GATGAAATTC GTGACCATTT 3240

ACGAGGGTTT GCAAGAACAA TTCCAAGTTT TATTATGGCT TACGGTGATC AAACTCTAAC 3300

ACTTGATAAT TTTGATGCCT TTGTTCCTGA ACATGTTTTT TATGAAGTAA CAGGGATTAC 3360

GATTGATCAG TTTAGATATT TGCGAGATGG TGGGCAGGAT TTTGCAGGGC ATCTCTTTGA 3420

TAAAGCAACA TTTGACGAAG CTATTCAAGA ATTTCTTCGC AAGAAAAAGG AGTTGGCGGA 3480

TTATTTTAAA GATCAAAAAG AAGACATTTT TGACTATATT CCACCGCAGA AGACCAACCA 3540

AATTTTCACT CCTAAACGAG TGGTGAAAAG GATGGTAGAT GATTTGGAAA AGGAAAATCC 3600

AGGGATTTTT GATGATCCAT CTAAGACTTT TATTGATTTA TATATGAAGT CAGGCCTCTA 3660

TATTGCAGAA CTTGTGAAGC GGTTATATAA TAGCAATGGC TTGAAAGAGG CCTTTCCAAA 3720

TCCTGAAGAA CGCTTAAAAC ATATTTTGGA AAAGCAAGTT TATGGATTTG CTCCGTCTGA 3780

GATTATCTAT AACATTTCCA CTAATTTTAT ATTTGGCAAT CTTTCTAAAG ATATCAGTAG 3840

GAAGAATTTT GTTTTAGCAG ATACCATTCC AGCGGCTAAA GAAGGGAGCA TTCAAAAGTT 3900

GGTTGATTCC TATTTTGAAA ATAATTAAAA AGAAGGCCGA GTCAAAATTC TTTGAAATCA 3960

GAAAAAACGC ATAATATTGA GTGCTTTTGT ACTGCCCCCC AAAAGTTAGA CAGAAAAAAT 4020

CTAACTTTTG GGGGGCAGTT CAGACAATCC TTGGTATTAT GCGTTTTATT GTGGGAAGAT 4080

GTATAATGGA TTGAAATAAG ATATGAACAA ATCAATTAGG AATTTAAAGC ATTTTATAAC 4140

AACGTTTTAG AGTAATGGGG GGCTATTTCA ACTTCAACCT ACTATAATAC AGAAAAAAAC 4200

AACTCCCTGA TAATTCAAGG AGTTGTCTAT AGTTAAATTA GTTTTTAGAA GCTTCTTGGA 4260

ATTCTGGGTT TTTCCATGCT TCGTCAATGA TAGCTTGTAA TTCTTTAGCA GATGCTTGCA 4320

TTTTTTGAGT TTCTGCGTCG TTCAATGGGA TATTTACTGG ACGAACGATA CCATGTGCAC 4380

CAACAACAGC TGGTTGACCG ATAAAGACAT TCTCAACTCC GTATTGACCT TCTTGGAATA 4440

CTGAAAGTGG AAGTACTGCG TTTTCATCGT CAAGGATTGC TTTAGTGATA CGAGCAAGGG 4500

CTACTGCGAT ACCGTAGTAT GTTGCACCTT TTTTGTTGAT GATTGTGTAG GCTGCATCAC 4560

GAACACCTTC GAACAATTCA ATCAATTCAG CTTCTTGAAC ATTTTGAGTG TCTTTAAGGA 4620

ATTCTTCAAG GTTTACACCA GCGATGTTAG CGTGTGACCA AACAGCGAAC TCAGAGTCAC 4680

CGTGTTCACC CATGATGTAG GCGTGCACTG AACGAGCATC CACATCCAAT TTTTCAGCAA 4740

GTGCTTGACG GAAACGAGCT GAGTCAAGTG AAGTACCTGA ACCGATAACG CGTTCTTTAG 4800

GGAAACCAGA GAATTTCCAA GTTGAGTAAG TCAAAACGTC AACTGGGTTA GCAGCAACAA 4860

GGAAGATACC TTTGAAACCA GATTCAACAA CTTGAGTTAC GATTGATTTG TTGATAGCAA 4920

GGTTTTTACC TACAAGGTCA AGACGAGTTT CACCTGGTTT TTGAGGTGCA CCTGCAGTGA 4980

TCACAACAAG GTCAGCGTCT GCACAGTCAG AGTATTGAGC TGCATAGATT TTTTTAGGTG 5040

AAGTGAAGGC AAGGGCGTGA CTAAGGTCAA GCGCATCACC AACAGCTTTT TCATGCAATT 5100

GTGGAATTTC GATAATTCCA AGCTCTTGTG CAATTCCTTG GTTAACAAGT GCAAAAGCGT 5160

AAGATGAACC TACAGCACCA TCACCGACAA GGATAACTTT TTTGTGTTGT TTAGTTGAAG 5220

TCATTGTTTT AAACATCTCC TTAATTTTAT TAGGGGATTT TCCCTAGACA ACTTCATT 5278

(2) INFORMATION FOR SEQ ID NO: 228:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1941 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

ATAAGGAATC TCTAAAAAAT TTTAAGGAGA ATCTAGCAAA TGGATTTCAC ATGGGCACTG 60

AAGTATGCCA CTGAATTTTT GGGAACTGCC ATTTTGATCA TTCTTGGGAA TGGTGCAGTT 120

GCCAACGTTG AACTTAAAGG TACGAAAGGT CACCAAAGTG GCTGGATCGT CATCGCTGTT 180

GGTTATGGTA TGGGGGTTAT GATCCCAGCC TTGATGTTTG GTAACGTATC TGGGAATCAC 240

ATCAACCCTG CTTTCACTCT AGGGCTTGCA GTTAGCGGTC TTTTCCCTTG GGCACAAGTG 300

GTACCTTACA TTATCGCGCA AGTCTTGGGG GCTATCTTTG GCCAAGCCTT AGTTGTGGCA 360

ACATACCGTC CATTCTACTT GAAAACTGAA AACCCAAATA ACATCTTGGG AACTTTCTCA 420

ACTATTTCAA GTATTGACCA TGGTACAAAA GAAAGTCGCT ATGCAGCAAC TGTCAATGGT 480

TTGATTAATG AGTTTGTTGG TTCATTTGTT TTGTTCTTTG CAGCTCTTGG TTTGACTAAA 540

AACTTCTTTG GTGCTGAAGT GCTTCAATTC ATGAAACAAA AGGCAACAGA AGCAGGACAA 600

ACAGTTGATT TTTCTGACTT GGCTATTAAA GCACAGGTGG CTCCACACAC TGCTTCAGGA 660

CTTTCTGTGG CTCACTTGGC ACTTGGATTC CTCGTTATGG CTTTGGTAAC ATCACTTGGA 720

GGACCTACAG GACCTGCCTT GAACCCAGCC CGTGACTTGG GACCACGTCT CCTTCATGCT 780

TTCCTTCCCA AATCAGTTCT TGGTGAGCAT AAAGGCGATT CAAAATGGTG GTATTCTTGG 840

GTACCAGTAG TAGCACCTAT CGCAGCAGCA ATTGCGGCAG TAGCTGTATT CAAATTCCTT 900

TATCTCTAAG AAATAGCTCC TTTAACATTT GAGTGAGCAC CATCTATAAG TAAGAGAGGA 960

TCAGACTGGk TCTCTCTTTT kGATTTTTaG GGAAATGAAA GAAcTCTAAA CAAACTCCTC 1020

TCCAGCAGTG GTTTAGAAGT CTCAGTGGGC TATTCCAGCT TCAATGGACT ATAGTAGGTT 1080

GCAGTTGAAA TAATAGACCC TTGTTTCTAA AACATTGTGA GAAATTGGTT TGAATTCTCC 1140

AATCAAATTG TGCAGTTTTC ATTCTACTAT ATATTATCGG AATATTATCG GAGATGGGTT 1200

CCCTATCTTG TAAGTCTGCT TTATAGTGGG TTGAAGTTGG AATAGTCCTC CCTTCTTTCT 1260

CAAACATTGT GAGGAATTGA TTTACCTTCC TCAACAAAAT GTTCAGTTTC TATTTCATTT 1320

TACTATAAAA TAAGCGATTA GGGGGGCTAT TCTTCGACCT ACATTGACTC TGCTGAGTCC 1380

TATGATTGTT ATCGTTTTAT CTGCAATTTT ATACTCAATG AAAATCAAAG GGCAAACTAA 1440

GAAGCTAGCC GCAGGTTGTT CAAAACACAG TTTTGAGGTT GTATAGTAGA TTGAAACTAG 1500

AATAGTACAC ATCTACTTCT AAAACATTGT TAGAAATCGA TTTGACTGTC CTGAACGATT 1560

TGCCCTATTC TTGTTTCATT TTACTATATA AACCAGAGAC TGTTTACATT TTCAGCAAGT 1620

GAGTGGATGG ATAATGCTGA AAACTCCTTG AAGGATAAGT CTATTTAGTA CTTTCTATTA 1680

ATTAGTTAAA TTTTTACCAA GAATAATTCA CAAAAACGTT GTAAAACACT TGCAATTTAG 1740

CTGAAATTTG ATAAAATAGT AAGGAAAGTT AGACTGTATT GCCTACTGTC TATCTATAAA 1800

ATATATTTTA TTGGAGGCTT TTACTCAAAT GGCAAAAGAA AAATACGATC GTAGTAAACC 1860

ACACGTTAAC ATTGGTACTA TCGGACACGT TGACCACGGT AAAACTACCC TAACTGCAGC 1920

TATCACAACT GTTTTGGCAC G 1941

(2) INFORMATION FOR SEQ ID NO: 229:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 755 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

ATTTGAAGAA ATTGAAGAAA TCGTAGCCCC TACAGATGGT GAATTTTTGG GGGAAGTTTT 60

ACTTGGAACT GGGGTAGTTC TCTTAATTGG AGTAGCCTGT TGTTAAAAAG ATAGGGAGTG 120

ATAATCATGC AAGATAACTT TTTATTTGAG GAAATTGAAG AAATTTCAGT ACCAGTTAAT 180

GATTTTTCAG CTGGACTTGC AACAGGTATC GGATTTGGTT TAGCAATCCT TGCTCTTGCT 240

GGTTGTTGAA GTTTGTTCAT TTACTAACAT CAAGCTTTTT CAATTTCATT TTAGACAGTC 300

ATTTAAATTT TCCGTATTAG TCTTGCAGCA AGAGATTAAT AGAATTAGTC ATTATTTTAT 360

TGATTGCGGA CTGAGGGACT AGAGTATGTT TTACTTAACC CCTCTTTTAT TTATTAAAGG 420

TTAGGTTTGT TATGAGAATT GTTGATAAGA TTAAGATATT ACCTACTCCT TATGAGGGAC 480

ACTATCATTT ATATATACCA TCCAGTAAGA AACATGTATT AGTTGGGAAA CAGGAAAAAA 540

ATGGTTAGAG CAACTAATAG GTCAAGAATT TACCATATCG GACTTATTAG TGTTAGTAGG 600

GAAGAAATAT TTTTAAAATA TCTTGGGACT TTAATATAAC ATTATCTGAA AAATTAAACT 660

ATAAAAGATT TAATAAGAAT TTTGAAAAAA TCCTATCTTG TTGTCATTAT ATTTGCAACG 720

ATACATGAAA TTAGTCATGC AATAATTGCT AATAA 755

(2) INFORMATION FOR SEQ ID NO: 230:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1483 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 230:

CCAGAAAAAC CGTAGTGGAG CTCGTGGAAC AGTGGAATTG ATTTTCCAAA AAGAATACAA 60

TAAATTTTCA AGTATCTCAA AGAGGGAGGC ATAAGATGTC AGATGCATTT ACAGATGTAG 120

CCAAGATGAA AAAAATCAAA GAAGAAATCA AGGCACATGA GGGACAAGTC GTAGAAATGA 180

CTTTGGAGAA TGGTCGTAAG CGCCAAAAAA ATAGATTGGG TAAGCTAATT GAAGTTTATC 240

CATCTCTATT TATTGTGGAG TTTGGGGATG TGGAAGGAGA TAAACAAGTT AATGTTTACG 300

TTGAATCCTT TACTTACTCA GATATTCTTA CAGAAAAGAA TTTGATTCAT TATCTTGACT 360

AAAGTGAGAA ATTTTCTCAC TTTTTCTTTT TTCTCCGAAT AATTTAGGTG AAGGCAATCA 420

TCGCTTTATA TTATTTTTCA AGGAGGAAGA ATGAAAATTT TACCGTTTAT AGCAAGAGGA 480

ACAAGTTATT ACTTGAAGAT GTCAGTTAAA AAGCTTGTTC CTTTTTTAGT AGTAGGATTG 540

ATGCTAGCAG CTGGTGATAG TGTCTATGCC TATTCCAGAG GAAATGGATC GATTGCGCGT 600

GGGGATGATT ATCCTGCTTA TTATAAAAAT GGGAGCCAGG AGATTGATCA GTGGCGCATG 660

TATTCTCGTC AGTGTACTTC TTTTGTAGCC TTTCGTTTGA GTAATGTCAA TGGTTTTGAA 720

ATTCCGGCAG CTTATGGAAA TGCGAATGAA TGGGGACATC GTGCTCGTCG GGAAGGTTAT 780

CGTGTAGATA ATACACCGAC GATTGGTTCC ATTACTTGGT CTACTGCAGG AACTTATGGT 840

CATGTTGCCT GGGTGTCAAA TGTAATGGGA GATCAGATTG AGATTGAGGA ATATAACTAT 900

GGTTATACAG AATCCTATAA TAAACGAGTT ATAAAAGCAA ACACGATGAC AGGATTTATT 960

CATTTTAAAG ATTTGGATGG TGGCAGTGTT GGGAATAGTC AATCCTCAAC TTCAACAGGC 1020

GGAACTCATT ATTTTAAGAC CAAGTCTGCT ATTAAAACTG AACCTCTAGC TAGCGGAACT 1080

GTGATTGATT ACTATTATCC TGGGGAGAAG GTTCATTATG ATCAGATACT TGAAAAAGAC 1140

GGCTATAAGT GGTTGAGTTA TACTGCCTAT AATGGAAGCT ATCGTTATGT TCAATTGGAG 1200

GCTGTGAATA AAAATCCTCT AGGTAAttCT GTTCTTTCTT CAACAGGTGG AACTCATTAT 1260

TTTAAGACCA AGTCTGCTAT CAAAACTGAA CCCCTAGTTA GTGCAACTGT GATTGATTAC 1320

TATTATCCTG GAGAGAAGGT TCATTATGAT CAAATTCTCG AAAAAGACGG CTACAAGTGG 1380

TTGAGTTATA CGGCTTATAA CGGAAGTCGT CGCTATATAC AGCTAGAGGG AGTGACTTCT 1440

TCACAAAATT ATCAGAATCA ATCAGGAAAC ATCTCTAGCT ATG 1483

(2) INFORMATION FOR SEQ ID NO: 231:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1027 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 231:

CCCGGAAAAC AAGTTAAAGT TGAAGTTGGT CAAGCAGTTT ACGTTGAAAA ATTGAACGTT 60

GAAGCTGGTC AAGAAGTTAC TTTTAACGAA TTGTTCTTGT TGGTGGTGAA AACACTGTTG 120

TCGGAACTCC ACTTGTTGCT GGAGCTACTG TAGTTGGAAC TGTTGAAAAA CAAGGAAAAC 180

AAAAGAAAGT GGTTACTTAC AAGTACAAAC CTAAAAAAGG TAGCCACCGT AAACAAGGTC 240

ACCGTCAACC ATATACAAAA GTTGTCATCA ACGCAATCAA CGCTTAATTT TAAGGAGAAC 300

ACATGATACA GGCAGTCTTT GAGAGAGCCG AAGATGGCGA GCTGAGGAGT GCGGAAATTA 360

CTGGACACGC CGAGAGTGGC GAATACGGCT TAGATGTCGT GTGTGCATCG GTTTCTACGC 420

TTGCCATTAA CTTTATCAAT TCTATTGAGA AATTTGCAGG CTATGAACCA ATCCTAGAAT 480

TAAACGAAGA TGAAGGTGGC TATCTGATGG TTGAAATACC AAAAGATCTT CCTTCACACC 540

AGAGAGAAAT GACCCAGTTA TTCTTTGAAT CATTTTTCTT AGGTATGGCA AACTTATCGG 600

AGAACTATTC TGAGTTCGTC CAAACCAGAG TTATCACAGA AAACTAACAC GGAGGAAAAC 660

ATTATGTTAA AAATGACTCT TAACAACTTG CAACTTTTCG CCCACAAAAA AGGTGGAGGT 720

1250

TCTACATCAA ACGGACGTGA TTCACAAGCA AAACGTCTTG GAGCTAAAGC AGCTGACGGA 780

CAAACTGTAA CAGGTGGATC AATCCTTTAC CGTCAACGTG GTACACACAT CTATCCAGGT 840

GTAAACGTTG GTCGTGGTGG AGATGATACT TTGTTCGCTA AAGTTGAAGG CGTAGTACGC 900

TTTGAACGTA AAGGACGCGA TAAAAAACAA GTGTCTGTTT ACCCAATCGC TAAATAAAAA 960

GGTCCATTGA ACCTTTTATC CCGAACCTTG AAATGTAGAG GTGAGGAAGC TAGAAACAGC 1020

TTAAAAT 1027

(2) INFORMATION FOR SEQ ID NO: 232:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1990 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

CGGTTCAAAT GGTGCAGGTA AATCTACGTT AATTAATTCT ATTGTAGGTT TTCAAGAGAT 60

TTATTTAGGA GAAATAGAGT ATTGTGATAA AGATTTGATA GTTAGTTCTC AACCTTTTGC 120

TCATTTAGGC TTTACTCCTC AAACCACAGT AATTGATTTT TATACTACTG TGAAGGACAA 180

TGTAATATTG GGGCTGAACC TTGCTGGAAA GTTTGGGAAA AATGCTGAGA AGTTGTGTCA 240

AATAGCCTTA GAAATTGTTG GGTTAGCTGA TAAAAAAAAT AATTTGGTAG AAACATTGTC 300

AGGTGGACAA CTGCAACGCG TCCAGATTGC TAGAGCAATA GCTCATAATC CAGATTTTTA 360

TATTTTAGAT GAACCTACCG TTGGTTTAGA TACTGAATCT GCCGAAAAAT TTTTAATGTA 420

TTTAAAAGAT AAGAGTTTGG AAGGAAAAAC TATTATCATA TCTTCACATG ACATAAATCT 480

ACTCGAAAAG TTTTGTAAAA AAATACTTTT TTTACAAAAT GGCTCCATAT CATTTTTTGG 540

TGATATGCGT GACTTTGTAG ATAATTCAAC TATCAAATTA AATTTTTCAA TGCAGAATAG 600

AATTTCTAGA TATCAAATTG AATTTTTAGA AAATTTTAGA TTTAAAGTTC ACATCGAAGA 660

TAATGATAGT TTTACAATAG AAGTCCCTAT AGAAGAAAAG ATCTTAGATG TTATCAATGA 720

GGTAGGAAAA GCATGTGAAA TTAAAAACTT TTCAACAAGT AAATTAACCT TACAAGAAAG 780

TTATTTGCAA AGAATAGGAG GAGAAAAATG AAGGCTGATC AATTAAGGCA CAAATCGGAC 840

TTAGGTTTAA GAGGTCTAGC GATTATTGCT AAAAATGAGA TTATTGCTTT TTTTAGAAGT 900

AAAGGTTTAA TTATTTCTCA GTTTCTACAA CCAATCTTAT ATGTTGTTTT TATAATAATA 960

GGATTAAATT CTTCGATAAA GAACATTCAG TTTAATGATA TAAAAACCTC TTATGCAGAA 1020

TATACAATCA TTGGTGTTAT AGCTTTATTG ATAATCGGGC AGATGACTCA AGTTATTTAT 1080

AGGGTGACAA TAGATAAAAA ATATGGGCTA CTTGCTCTTA AGTTATGCAG TGGAGTTCGT 1140

CCTTTATATT ATATTTTAGG GATGAGTATC TATTCTATAT TAGGGTTGAT AGTTCAAGAA 1200

ATTATTATAT ATATAATTAC GTTAGCGTTT GAGATAAATA TCGCAATGGA TAGATTTTTT 1260

TATACAGTTT TGTTATCTAT TGTTGTTTTA TTATTTTGGG ACTCCCTTGC AATTTTACTT 1320

ACAATGTTTA TCAATGATTA CAGAAGACGT GATATTGTAA TACGTTTTGT ACTAACACCG 1380

CTTGGTTTTA CAGCTCCTGT TTTCTACTTA ATAGATTCTG CTCCTAGTAT TGTGAGATGG 1440

ATTGGTCAGT TAAATCCCTT AACTTATCAA TTAACTATTT TGAGAAACTT TTATTTTAAA 1500

AATTCAACAA CTTTGGAATT AGTTTTCTTA TTGTTAACAT CATTACTTGT CCTTATATCT 1560

GTATCTTTTA TTATACCAAA GATAAAATTG ATACTGATAG AAAGATAAAA GTTGGGTCAT 1620

CCAACTTTTT TGTTGTCTCC CGAAAACCAC TAGCTATGCT AGTGGTTCCA TAGAGCTTTT 1680

AGCGTGGTAA CAAAAAGAAC CTCCTAAAAT GATAAGATAG AAGTGGTTTC TCCGCCACTA 1740

CAACATATCA TACAGGAGGT ACCTCATGAG AGAGGATAAT CAAAGTTTAT CACATACCAC 1800

ATGGAATTGT AAATATCATA TTGTTTTTGC ACCCAAATAT CGTCGTCAAA TCATTTATGG 1860

CAGATACAAA GCTAGTATCG GAAGAATCAT ACGTGACTTA TGTGAGCGTA AGGGTGTAAT 1920

AATCCATGAA GCGAATGCTT GTTCAGACCA TATTCACATG CTTATCAGTA TTCCTCCGAA 1980

ACTTAGTGTT 1990

(2) INFORMATION FOR SEQ ID NO: 233:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4766 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

GAACTATATT GCATATATTT CTAGCAATGA TCATGGCGAA TCTTGGTCTG CACCAACTTT 60

ATTACCTCCT ATAATGGGAC TTAATCGGAA TGCGCCATAT TTAGGTCCTG GACGTGGAAT 120

CATTGAAAGC TCAACTGGAC GTATTCTTAT TCCGTCTTAC ACTGGTAAAG AGTCTGCGTT 180

CATTTATAGT GACGATAATG GAGCATCTTG GAAAGTTAAA GTAGTGCCAC TTCCTTCTAG 240

TTGGTCAGCA GAAGCACAAT TTGTAGAATT GAGTCCAGGA GTAATTCAAG CATATATGCG 300

TACAAATAAT GGTAAAATTG CATATTTAAC AAGTAAAGAC GCAGGTACTA CTTGGAGTGC 360

ACCGGAATAT TTGAAATTTG TTTCAAATCC AAGTTATGGA ACACAATTAT CAATCATCAA 420

TTATAGCCAA TTGATTGATG GTAAAAAGGC TGTCATTTTA AGTACTCCAA ACTCCACAAA 480

TGGTCGTAAA CACGGACAAA TTTGGATTGG TCTAATTAAT GATGATAATA CAATTGATTG 540

GCGTTATCAT CACGACGTTG ATTATAGTAA CTATGGATAC TCATATTCAA CATTGACAGA 600

GTTACCAAAT CATGAAATTG GATTGATGTT TGAAAAATTT GATTCATGGT CTCGTAATGA 660

ACTTCATATG AAAAATGTTG TACCATATAT AACATTTAAG ATTGAAGATC TGAAAAAGAA 720

TTAAAGCTGA AATTTGAAAA TATATAAAAA GAGGATAAAA ATTATGGTAA ATTACGGTAT 780

TGTTGGAGCT GGATATTTTG GAGCTGATTT AGCTCGCTCA ATGAACAAAA TTGAAGATGC 840

AAAAGTGGTT GCGGTATTTG ACCCAAATCA TGGAGAAGAA GTTGCTCAAG AGTTGGGATC 900

AGATGTTTGT GCAAGTTTAG ATGAACTTGT AGCACGTGAA GATATTGATT GTGTGATCGT 960

AGCTTCACCT AGCTACCTTC ACCGTGAACC AGTTGTGAAA GCTGCTCAAC ATGGCAAACA 1020

CGTATTTTGT GAAAAGCCAA TTGCATTGTC TTATGAAGAT TGTAAAGCCA TGGTTGACGC 1080

ATGTAAAGAA AATAATGTCA TCTTTATGGC TGGTCACATC ATGAACTTCT TTAACGGTGT 1140

ACACCATGCT AAAGAATTGA TTACTCAAGG TAAAATCGGT AAAGTTCTTT ATTGCCATGC 1200

TGCTCGTACA GGTTGGGAAG AACAACAACC AACTGTATCA TGGAAGAAAC TTCGTTCTCA 1260

ATCTGGAGGA CATTTGTACC ACCATATTCA TGAATTAGAT TGCATTCAGT TTATCATGGG 1320

AGGACTTCCT GAAAAAGCGA CAATGGTAGG AGGCAATGTA TATCATAAAG GTGAAAACTT 1380

TGGTGATGAA GATGATATGC TCATTGTAAA CTTAGAATAC TCTGATGATC GTTATGCTGT 1440

TTTGGAATAT GGTAATGCTT TCCGTTGGGG TGAACACTAC GTCTTGATTC AAGGAACTGA 1500

AGGAGCTATC AAACTTGACT TGTTCAATAC TGGCGGTACT CTTCGTGTTA AAGGTGAAGG 1560

AGAATCACAC TTCTTAGTTC ATGAAACTCA AGAGGAAGAT GATGATCGTA CAGCTATCTA 1620

TACCGGTCGT GGTATGGATG GAGCAATTGC GTACGGTAAA CCAGGAGTAC GTTGCCCATT 1680

ATGGTTGCAA ACATGTATTG ATAAAGAAAT GGAATATCTA CATGACATCA TTAAAGGTGG 1740

AGAAATTACA GAAGAATTTG AAAAACTTCT CAATGGTGTA GCTGCTTTAG AATCAATCGC 1800

TACCGCTGAT GCATGTACTT TATCAGTTAA AGAAGATCGA AAAGTAAGTC TTTCAGAAAT 1860

CACAAATGCT TAACTTTTGT AAAACAGAAT AGTAAATTCT TGTCATTATA TAATTTCTAA 1920

AGTTCTGTGA TACAACTCAT TGAATAAAGA AATAGAGATG GGACTGGGAT AATGCCCAGT 1980

CCCATTTTTT ATCAAAAAGT AATGAGATCA AAAATGTGGG AGTGTTGAAA TGAAGATTAT 2040

AGGTATCGAT ATTGGCGGAA CAACAATTAA GGCAGATTTA TACGATGAGT TTGGAACGAG 2100

TTTGAATCAT TTCAAAGAGA TAGAAACAAT TATTGACTAT GATTTGGGAA CGAATCAGAT 2160

ATTAAATCAG GTCTGTGATT TAATTGGTGA GTATACTTTA AATCATTCAA TTGATGGTGT 2220

TGGGATTTCC ACTGCTGGAG TTGTTAATGC TAATACTGGA GAAATCATCT ATGCAGGCTA 2280

TACAATACCA GGGTATATCG GAGTAAACTT TACTGCCGAA ATAGAAAAAC GTTTTGGGTT 2340

GTATACTTTT GTTGAAAATG ATGTTAATTG TGCTGCATTA GGTGAATTGT GGAAGGGACA 2400

AGCCAAAGAT AAGAAAAATG TAGTAATGGT TACTATTGGA ACAGGTATAG GAGGCAGTAT 2460

TATTGTCAAC GGACAAATTG TTAACGGATT TAACTATACT GCTGGTGAAG TAGGTTATAT 2520

TCCTGTAGGT AATTCGGATT GGCAAAGTAA AGCCTCAACA ACCGCATTGA TTCATTTATA 2580

TCAAAAAAAG AGCTTGAAAA CTAATCAAAC TGGACGTACT TTCTTCACTG ATTTAAGATC 2640

TGGAGATAAA GTTGCTGAAG AAACTTTTGA AATTTTTGTA GAAAATCTAA CAAAAGGTTT 2700

ATTAACGATT TCTTATCTAC TTAATCCAGA AATTCTCATA TTAGGAGGTG GGATTCTGGA 2760

TAGTAAGGAT ATTTTGTTAC CTGAAATTCA AAGTTCTTTA GCTAAAAATG CAATGGATAA 2820

TAGGTTTTTA CCTAAAAATC TTGTGGCAGC TACATTAGGA AATGAAGCTG GTCGTATAGG 2880

AGCTGTAAAA AATTTCTTAG ATAGAATTTC TAATAAATAG TATGTAAGAT AAGGAGGTGT 2940

CACAATGACT AACTCTGTAT TTTCGACAAT GCAAGATATT GAGAATGTTG CAACCGATAT 3000

TATAAAATCA TATGATAATG AGATTTATAC TTATAAAGCT GTTTCCCAAG AAGAATTGGA 3060

AAAACTAGAA AAAAGTTATG ATGAAAAAAG TCACGAAGAA TTAGTTTCAA TAGAAAGCAA 3120

TTTAGAAATG AAACAACAGA ACCTTATTGA TGAGGTTAAT AAAACAATCA AGGAAAATGA 3180

TGCAAATATT CAGTATATTT CATCAAGTAG GAGAGGAGAA TTTGTAGAAA AAATTATTGG 3240

TAGGGTGGTA GAAAAATATG GCCATTAGTC AGATGAAAAG AATCTCTCTA CTATTTTCTA 3300

AAAGTAGTCT TGATGATGTT TTAAAAACTA TTCAAGAACT AGAGTCAGTG CAGTTCCGTG 3360

ATTTAAAGGT TCAGGATAAC TGGTCAGAAG CTCTAGAAAA AGATGAAGTT GTATTTCCAA 3420

CTATTCAAAT TTTTCATACT TCTAATTCCA ATCATGGGGT TATTGAGGGA AATGATGCCT 3480

TGACTTATTT GATGAATCAA CAACAACATT TAGAAGCAAC TGTAGAGAAA TTACAAGAAT 3540

ACCTACCGAA AGAAAACACG TTTAAATTAT TGCAGCAACC TCCGATAACT ACCTCTTATG 3600

AAGAATTAGA GAAATTTGGT AAAGCTAATG TTGCTGAGGG TGTTCTTAAA AAAGTGAATC 3660

ATCAAATTAA CAGAGTTCAT GAATTAGAAA GACACATTCA AAGTAATAAT GAGGAAATAG 3720

AGCGATTAAT AAAGTGGGAA AAATTAGAAA TTGTTCCTGC GAATTTAGAA CAATTTTCTT 3780

TCTGTAAAGG AAAAGTCGGA ACAATTCCAA GGACTGAAGA TAATCGCTTA TACAATAGTC 3840

TTTTAGAAAA CAATATTGAA GTTCAAGAAA TATTTTCTAA TGATAGAGAG TACGGTGTTG 3900

TTGTTTTCTA TCAGTCTAGT TACTCTATAG ATTTTGATGA ATACTTATTT GAACCATTTG 3960

ATTATTCTAG AAAGGAATTA CCGAAGCAGC GAGTAGTAGA TTTAGATCAA GAAAACATGC 4020

AGTTAATAAC TGAAAAAGAG AATATTATCG CATCGTTGCA AGATTCAAAG AAATATTTGA 4080

TAGATTTACA ATGGCAAATA GACTATATTT TATCTATCTA TGCTCGTCAA ATCTCTAAGA 4140

ATAACTTTTT GTGCACTCCG CATCTAGTTG CATTAGAAGG ATGGATAGAA GAAACTCGTA 4200

TTTTATATTT TATAAAAGTT ATGGATGAGC ATTTTGGACA TTCTATTTAT ATTTATGAAT 4260

CGGAAACATT GACGGATAAT CAAGATGAAA TACCTATCAA ATTAACGAAT CATTCTTTAA 4320

TTGAACCATT TGAATTATTG ACAGAAATGT ATGCTCTGCC CAAATATTAT GAGAAAGATC 4380

CTACACCTGT ATTAGCACCA TTTTACTTTA CATTTTTTGG AATGATGGTT GCTGATTTAG 4440

GCTATGGTTT ACTATTGTTT TTAGGAACAA TGTTAGCATT AAAAATTTTT CATCTACCTT 4500

CAGCAACTAA GAGATTTTTA AAATTCTTTA ATATATTAGG GGTAGCCGTT GCAATTTGGG 4560

GTGGAATCTA TGGCTCATTT TTTGGATATG AGTTGCCATT TCATCTGATA TCTACAACCT 4620

CTGATGTCAT GACTATATTA GTAGTGTCAG TTGTGTTTGG GTTTATTACA GTATTTGCAG 4680

GTTTGTTAGC TTCAGGACTA CAAAAAGTAA GAATGAATAA ATATGCAGAA GCATATAATT 4740

CAGGATTTGC GTGGTGTGTT ATTCTG 4766

(2) INFORMATION FOR SEQ ID NO: 234:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2484 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 234:

CCTTTTAGAA AAAATTAAAG AATACGACAC CATTATCATT CATCGTCATA TGAAACCAGA 60

CCCTGATGCC TTGGGAAGTC AGGTGGGATT GAAAGCCTTG CTGGAACATC ATTTCCCAGA 120

AAAAACCATC AAAGCCGTCG GTTTTGATGA ACCAACTCTT ACTTGGATGG CTGAGATGGA 180

TCTTGTTGAA GATAGAGCCT ACCAAGGCGC ACTTGTCATC GTCTGTGATA CAGCTAATAC 240

TGCTCGTATC GATGATAAGC GCTATAGTCA AGGTGATTTT CTCATTAAGA TTGACCACCA 300

TCCAAATGAT GATGTATACG GTGACCTGTC TTGGGTCGAT ACTAGTTCAA GTAGCGCTAg 360

aGaTGATTAC CCTATTTGCC CAAACAACCC AACTAGCCTT GGCAGATCGC GATGCTGAGT 420

TGCTCTTTGC AGGAATTGTC GGTGATACAG GTCGCTTCCT CTACCCTTCT ACCACTGCAC 480

GGACTCTTCG CCTGGCTGCT TATTTGAGAG AACATAACTT TGACTTTGCG GCTCTCACTC 540

GCAAAATGGA CACTATGAGC TACAAAATTG CTAAACTGCA AGGCTACATC TACGACCATC 600

TGGAAGTGGA TGAAAATGGT GCTGCTCGCG TTATCCTGAG TCAGAAAATC TTGAAACAAT 660

ACAATATAAC CGATGCTGAA ACTGCGGCCA TTGTAGGTGC ACCTGGACGC ATTGACAGAG 720

TGAGTCTCTG GGGAATTTTT GTCGAACAGG CTGATGGCCA CTACCGAGTT CGCTTACGCA 780

GTAAAGTCCA TCCTATCAAT GAAATTGCCA AGGAGCATGA TGGTGGAGGC CACCCTCTAG 840

CAAGTGGTGC TAATTCCTAT AGCCTAGAAG AAAACGAAAT CATCTACCAA AAGTTAGAAG 900

ACTTGCTTAA AAACTGATAA AATACTTGCC AAACTTTTCA GAATCTGATA GACTAGTATA 960

GTAACAATCT ATGGCTCGCA AAGAGACCAT GGCAGAAAGG AAATATTGCA AAATGAAAAr 1020

AGATATCCAT CCAGAATATC GCCCAGTTGT CTTCATGGAC ACAACTACTG GTTACCArTT 1080

CCTTAGCGGT TCAACAAAAC GCTCTAACGA AACAGTTGAG TTCGAAGGCG AAACTTACCC 1140

ATTGATCCGT GTGGAAATTT CATCAGACTC ACACCCATTC TACACTGGAC GTCAAAAGTT 1200

CACTCAAGCA GATGGACGCG TGGATCGTTT CAACAAAAAA TACGGTCTCA AATAATGATA 1260

AGAGAACAGT TTTGGCTGTT CTTTTTTGTT TCTTGAAATC AACTGCTGTT TTCATGTTCC 1320

AGACTCATCT GTAGGTTCGA TTTCCATGCT ACTAGGCAGG AAGGAAATAG CTGTTTCAAC 1380

ACGTCCATAA TGAGCTATAC TATTGTCACG AACCACACTT TCATTGATGG TCCAAGTGGA 1440

ATTCATTTTC TTAAAAGCTT CTCGGACTTT TTCCAAATCT TTGGAGGCAA TGGCCTGCTC 1500

TAAGGTTTCA AAACGAGGAC TTATACTCAT CTGCTTTCAA AAAGCATTCT AGTCCATCTC 1560

CGATTACCGA TGGACTTTAT CACCTCCTTC TCCAGTCCTT GTATGACATC TTGAAGTTGA 1620

TTCATGACAT CTTCCAAAGT TCgAAAGGCT TTATTCTTAA ATCCACGTTT ACGAATCTCT 1680

TTCCACACTT GTTCAATGGG TTCATCTCTG GTGTGTATGG AGGAATAAAG GTAAAATCAA 1740

TATTAGTCGG AATATTTAAG GTACTTGATT TATGCCATAT AGCATTGTCC ATAACGAGTA 1800

AAAGGATAAG CTTGTGAAAG CTCTTCTAAA AAGGCGTTCA TCCACACTCC TTTTTATAAA 1860

CCTGAAATAA GGCATCAATT GTAACAAATT CTCCTGCCTC TGTAGCCTTC AAATGACGGG 1920

CAAGAAAGGC TTTCTCTTCC TCAACTGTCA TATATGCATG GTTACGACCA CCACGTGTTT 1980

CTTGAAGGAG AGAGTCGAGT CCGAACTCCT CATATTTTTT TACGTTTCGC CAAATCGTTG 2040

TTTGATTACA GTCTAAAAGC TCTATAATCT CTTTATAAGA TTTGCCCATC AGACGAAATA 2100

TAGTAGATTG AAACTAGAAT AGTACACCTC TACTTCTAAA ACATTGTTAG AAATCGATTT 2160

GTCCTGTTCT TGTTTCATTT TACTATAGAA CGATTTGAAG GCGTTTATAA TATTTAGCTG 2220

TACGAGAGTC TTTTAAAAGT GTTTTGATGG TTTGGATTTC TTCTTTAGTT GATTTCATAT 2280

TACTATTATA TAATGCTTTT TGATTTTAGT CTGGTATAAA TATTGCTTTC CTCCAAAATG 2340

GTCATAGTTT TACTGGCAAA TCTAACATAT CACGGATAAA TTAACAAGTG ATTTCTGAAT 2400

TGCTAAACAT TTTCTTTTCT TATAGCATAC TTTAAGATTT TGTCTTTGAG AAAGATATTT 2460

CCAAGAAAAA CGTTCGTTTT TTGG 2484

(2) INFORMATION FOR SEQ ID NO: 235:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1766 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 235:

CTAGATATAG CTATAATTTT ATTTATAACA AGAGGATAGA AATGACCGAA TTAGAAAGAA 60

AAAATCGAAA AATTAGCTAA GAAATATTCT GATAACTTAA ACATCAAAGT TCAAGAGAGA 120

GTTCGTGAAA TGGCAAATGA TAATAAGAGC CATTATTTGA TATACAGAGT TTTAGGTATT 180

TCATTTGAAG AAGGAGAAAA TATCGATTTG TATCAAAATA AAGGTCGTTT TTTATACAAA 240

TATGCTGGTT CATTTTTAGA AGAAGCTGCA GTACTATGCT TTAACGAAAA ATTTGGTACA 300

GAAAATACTT AAAAAGTTAA CATTCCTAAT TCTGAAAGTA CAAAACCTAA GACTTTTGAA 360

ATTGATTGTT TAGTCGGAGA AAAACACGCA TACGAAATAA AATGGTGGGA TGCAACTACA 420

GATGGAGACC ATATAACTAA AGAACACACT AGAATAAAAG TTATTCATAA CAAAGGATAT 480

ATACCAATTC GGTTAATGTT CTACTATCCA AATAGAACTC AAGCTATAAA AATTCAGCAA 540

ACTTTAGAAA CATTGTATAA CGGTATTGGA GGGAAATATT ATTATGGAGA TTCTGCCTGG 600

GAACATTTAA GAGCAGTGAC CGGTATTGAT TTACTTAGTA TTCTAACAGA TATTGCAAAT 660

AAAAAAACAG GGGTAAAATC AAAATGACAG TATTAAAAGG AGATAACTTA GAAATATTAA 720

AAACTATTGA ATCCTCAAGT ATTGATTTAA TCTATATGGA CCCTCCTTTC TTTACACAGA 780

AAACCCAAAA ATTATCTAAT AACAAAAATA TTATGTATTC ATTCGAAGAT ACGTGGACAT 840

CGATTGAGGA TTACAAAGAA TTTTTGTCTG TAAGATTAGA AGAATGCAAA AGAGTGCTAA 900

AAAATAGTGG CAGTATTTTC GTTCATTGTG ATAAAATTGC AAATCATCAT ATTAGATTAA 960

TTTTAGATAA TATCTTTGGA GTAGATATGT TTCAAAGCGA AATTATATGG AACTATAAAC 1020

GGTGGTCTAA TTCAAAAAAG GGATTATTGA ACAATCATCA AAACATTTAC TTTTATTCAA 1080

AGTCAAAAGA TTTTAAATTT AATACAATTT TTACAGAGTA TTCTTCTACT ACAAATATCG 1140

ACCAAATACT AGTGGAACGA AAACGAGATG GAAACTCTAA AACTATATAT AAGGTTGATA 1200

ATAATGGTAA CTATATTCTA GCAAAAGAGA AAAATGGAGT TCCCCTTTCA GATGTTTGGA 1260

ATATACCATT TCTTAATCCA AAAGCTAAAG AAAGAGTAGG TTATCCTACA CAAAAACCTA 1320

TTCTGTTATT AGAACAAATT ATAAAGATTG CTACTGATAA AAATGATATA GTTTTAGACC 1380

CGTTCTGTGG AAGTGGAACT ACTTTAGTAG CCTCCAAGAT TTTGAATAGA AATTATATGG 1440

GGATTGATTT ATCTGAGGAA GCTATCAATA TAACTCAGCA ACGTCTGGAA AATGTTATAA 1500

AAACAAGTTC AAATTTATTG AATAAAGGAA TCGAAGCATA TAGAACCAAA ACTGAGGAAG 1560

AGGAAAACAT TCTTAAATTA TTACAGGCAA AAATTGTTCA AAGAAATAAA GGAATTGATG 1620

GTTTTTTACC TAAACATTTT CAAAAAAAAC CGATACCTAT AAAAATTCAA AAAAATAATG 1680

AATGTCTGAA TGAGAGTATC TCTTTATTAC AGAATGCTAT AAACTCCAAA AAACTTGATT 1740

TTGGAGTAGT TATAAAAACT CATTCG 1766

(2) INFORMATION FOR SEQ ID NO: 236:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 748 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 236:

CCGAAAATCA AATTCAAACC ACGTCAACGT CGCCTTGCCG TACTCAAGTA CAGCCTGCGG 60

CTAGTTTCCT AGTTTGCTCT TTGATTTTCA TTGAGTATTA AACTAAATTA AATAATATTA 120

GCGCGGAGAA TTTCTAATTC TTCCTTGGTC AAGCGACGCC ATTCCCCTCG TTCTAGGTTC 180

TCATCTAATA CTAAAGTTCC CATAGTCAAT CGTTGCAAGT CCACCACTTC CTTGCCACAG 240

TAGCCCACCA TACGCTTGAT CTGATGAAAC TTCCCTTCTG CAATGGTCAC ACGGATTTGG 300

CTTTGATTCT TTTCTGTATC TATGGATACA AGCTCCAGTA TAGCGGGTTG ACAGGTAAAG 360

TCTTTGAGAG GAATACCCTC AGCAAATGTC TCCACATCTT CTTGGGTCAT GATTCCCTTG 420

ACTTGTGCCA GATAAGTCTT GTCCACATGA CGCTTGGGCG AAAGAAGAAC ATGAGCCAGC 480

TGACCATCAT TGGTCAAGAG CAAAAGACCA TGCGTGTCAA TATCCAAGCG TCCTACTGGG 540

AAAACTTCCT TACTCCGCGC CAAGTCATCC AACAAGTCCA GAACGGTTCT GTGCTTGGGA 600

TCCTCAGTCG CTGAGATAAC TCCTTTGGGC TTGTTCATCA TGTAGTAGAC AAACTCTTCA 660

TACTCCAACA CTTGCCCATC AAAGCGAATC TCATCTATTT TTTCATCAAT CTGCAATTTA 720

GCTGATTTTT CTTTTTGACC ATTTACAG 748

(2) INFORMATION FOR SEQ ID NO: 237:

1258

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1449 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

AAAAGATTAC ATTGCAACAA TTGAAAATTA TCCAAAGGAA GGCATTACCT TCCGTGATAT 60

TAGTCCTTTG ATGGCTGATG GAAATGCTTA TAGCTACGCT GTTCGTGAAA TCGTTCAGTA 120

TGCTACTGAC AAGAAAGTCG ACATGATCGT GGGACCTGAA GCTCGTGGAT TTATCGTGGG 180

TTGTCCAGTT GCCTTTGAGT TGGGAATTGG TTTTGCGCCT GTTCGTAAGC CAGGTAAATT 240

GCCACGCGAA GTTATTTCTG CTGACTATGA AAAAGAGTAC GGTGTCGATA CCTTGACTAT 300

GCACGCGGAT GCCATTAAGC CAGGTCAACG TGTTCTTATT GTAGATGACC TTTTGGCGAC 360

AGGTGGAACT GTTAAGGCAA CTATCGAGAT GATTGAAAAA CTTGGTGGTG TTATGGCAGG 420

TTGTGCCTTC CTTGTTGAAT TGGATGAATT GAACGGCCGT GAAAAAATTG GTGACTACGA 480

CTACAAAGTT CTTATGCATT ATTAATGAAA ACAGTCCCTA GGGCTGTTTT CTCTACACTA 540

GGATATAAAA ATAGACTATA ACTAGTTAGA GAAAAACTAT AATTGAAAAC TATATCTTCT 600

TGCAGTATAA TAAAAGGACT AAGTGTTTGA GATTTGTCTT CAAACATATG CAATTATTCC 660

TGAAAGAGTA CAGTTAGGAG AGGGTTATGC CGATTCGAAT TGATAAAAAA TTGCCAGCTG 720

TTGAGATTTT ACGGACAGAG AATATCTTTG TCATGGATGA TCAACGTGCT GCCCACCAAG 780

ATATCCGTCC TTTGAAGATT TTAATTTTAA ATCTCATGCC ACAGAAAATG GTCACAGAGA 840

CCCAGTTGTT GCGCCACTTG GCTAATACAC CCCTACAACT GGATATTGAT TTTCTCTATA 900

TGGAGAGCCA CCGTTCTAAA ACAACTCGTT CAGAGCACAT GGAGACCTTC TATAAAACTT 960

TTCCTGAAGT CAAGGATGAG TATTTTGATG GGATGATCAT CACGGGTGCT CCAGTTGAGC 1020

ATTTACCATT TGAGGAAGTG GACTATTGGG AGGAATTTAG ACAGATGCTT GAGTGGTCTA 1080

AGACTCATGT CTATTCGACC CTTCATATCT GTTGGGGGGC TCAGGCTGGG CTTTATCTGC 1140

GCTATGGTGT AGAAAAATAC CAGATGGACA GTAAGCTATC AGGTATTTAT CCTCAGGACA 1200

CCCTAAAAGA GGGTCACCTT CTATTTAGAG GCTTTGATGA TAGCTATGTA TCCCCTCATT 1260

CACGGCACAC GGAGATTTCT AAGGAAGAGG TCTTAAACAA GACCAATCTC GAGATTTTAT 1320

CAGAAGGACC TCAGGTTGGG GTTTCTATTw TGGCCAGTCG TGATTTACGA GAAATTTATA 1380

GTTTTGGTCA TTTGGAGTAT GACCGTGATA CTTTGGCAAA AGAGTATTTT CGAGATCGTG 1440

ATGCAGGTT 1449

(2) INFORMATION FOR SEQ ID NO: 238:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 904 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

TACCCGCTTC TTTCAAGAGT TGGAGCAGGG CTTGTTTGCG ATCTTTTGTC ATAGTTCTTC 60

CTTTTAACGG CGTTTTCGAA GCACTTTATA GACAGCTAGT GCTAATGTAT AGTCTACCAT 120

ACTATGGATA ATTGTACCAA ATCCAACTAG TACAAATAGA ACATAAAACA TATTTTCTAC 180

ATTGGTACCA GAAGTTGCGT AAAAAACGAC ACAGGCCAAT ACTTCAGCAA GGGCATGAAC 240

AACAGCCAAA ACAAAGTTGA AAATCCAGGA AGATTTTGGT TTATCTAGGG TATCGGGGAA 300

TTTTTGTAGG TAAAGAGCTC CTAAAGCACC AAAAGATATA TGGGAAAAAG CCCGAAAAAC 360

GATAACCATG GGATAGCCAG CCATCAAAAA TCCAAAACTA GAGGCTAGGA TGACAAAAAC 420

TGCCATCAAG GGCGACAAGA ACATGGCTAT AAAAATAGCG ATGTGGCTCC CCAAAGTATA 480

GGAAGCAGGT GGAATGACAA TCTTGAAAGG CATAACAATT GGAATCAAAA TCGCAATAGC 540

CGTTAAAAGG GCTGTCATTG TCATAAATTG TGTCTTTTTC CGTGTATTCA CAAGAATCTC 600

CTTTTTAACT GCATATACAC TAGTATGGTA CAATAAACCA GACAATAAAG CAAGAATTTA 660

CTTGGGTTTA TAGATCATTT TTTAGTTAAA AGTTATAGTA GATTGAAACT AGAATAGTCC 720

ACCTCTACTT CTAAAACATT GTTAGAAATC GATTTGGCTG TCCTGATCGA TTTGTCCTGT 780

TCTTATTTCG TTTTACTATA GTAAAGATTT CATTAAAAAG AAACTGTATA GAGCAAAATC 840

TCCACCTTCA GGTTTGGAAA GCGGAGATTG TTTnTTATTT TTTCCAGGGT TTGTAGTCGT 900

GGGA 904

(2) INFORMATION FOR SEQ ID NO: 239:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 946 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

CACTCAAACA TGACTTATAT CAAGACGGAT GGACTTCAAG ACGATGCCAA TCGCTTGAAT 60

1260

CGTAACATTC AGTTTGGTGT TCGTGAATTT GCAATGGGAA CAATCTTGAA CGGGATGGCC 120

CTTCATGGTG GACtTCGTGT ATACGGTGGA ACTTTCTTCG TCTTCTCTGA CTATGTGAAG 180

GCAGCTGTCC GCTTGTCAGC CTTACAAGGA CTTCCTGTGA CTTATGTCTT TACCCATGAT 240

TCAATCGCAG TTGGGGAAGA TGGTCCGACT CATGAACCAG TTGAGCATTT AGCAGGTCTT 300

CGTGCTATGC CAAATCTAAA TGTTTTCCGT CCAGCAGATG CGCGTGAAAC GCAAGCAGCT 360

TGGTACCTTG CAGTGACAAG TGAGAAAACA CCAACTGCCC TTGTCTTGAC ACGTCAAAAT 420

TTGACTGTTG AAGATGGAAC AGACTTCGAC AAGGTTGCTA AAGGTGCTTA TGTTGTATAT 480

GAAAATGCAG CCGACTTTGA TACCATCTTG ATTGCGACAG GTTCAGAGGT TAATCTTGCT 540

GTCTCAGCTG CTAAAGAATT GGCTAGTCAA GGCGAAAAAA TCCGCGTAGT CAGCATGCCA 600

TCTACAGATG TCTTTGATAA ACAAGATGCA GCTTACAAGG AAGAAATTCT TCCAAATGCA 660

GTCCGCCGTC GTGTTGCAGT CGAAATGGGT GCAAGTCAAA ACTGGTACAA ATATGTTGGT 720

CTCGATGGTG CCGTTCTAGG TATTGATACT TCGGAGCCTC TGCCCCAGCA CCAAAAGTAT 780

TGGCAGAATA TGGCTTTACT GTAGAAAATC TTGTAAAAGT TGTTCGAAAC TTGAAATAAT 840

CCTAAAAATC AGGGCGTAAG CTCTGGTTTT TCTTACCAGA AAAGTAAGGT ACAATCTTGT 900

AAAAGTAGCT GAAATTTGAT ATAGTAGTCC TATGTAAAAG ACAAAG 946

(2) INFORMATION FOR SEQ ID NO: 240:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2764 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

CGGGGCTCCc TAGTTCTTAG GGAGCTATTT TTGTTTTTTC AAGAAGTTAT CTTCTTGTAT 60

TTTATACTCA ATGAAAATCA AAGAGCAAGC TAGGAAACTA GCCGTAssTG CTCAAAACAC 120

TGTTTTGAGG TTGTAGATAA GACTGACAAA GTCAGGAACA CATATCTACG GCAAGGCGAC 180

GTTGACGCGG TTTGAAGAGA TTTTCGAAGA GTATTAGTTG TGAATCTGGT GCAGTCGTCC 240

CAGATTATTC TTATTAGTAG GGTCTTGTTT TCTATATCCC CTCGTAGTTA ACAAGACCTT 300

GAGCATTTTA GAAAGAGGAA TCTATGTCTA CGAAATATAT TTTTGTAACT GGTGGTGTGG 360

TATCGTCCAT TGGGAAAGGG ATTGTGGCAG CGAGTCTAGG CCGTCTCTTG AAAAATCGTG 420

GTCTCAAAGT AACCATTCAA AAGTTTGACC CTTATATCAA TATTGATCCG GGAACCATGA 480

GTCCTTACCA GCACGGGGAA GTTTTTGTGA CAGATGACGG AGCTGAGACA GATTTGGACT 540

TGGGTCACTA TGAACGTTTC ATCGATATCA ATCTCAACAA ATATTCCAAC GTGACAACTG 600

GGAAAATTTA CAGTGAAGTT CTTCGTAAAG AACGCCGTGG AGAATACCTT GGGGCAACTG 660

TTCAAGTCAT TCCTCATATC ACAGATGCTT TGAAAGAAAA AATCAAGCGT GCCGCTCTAA 720

CGACCGACTC TGATGTCATT ATCACAGAGG TTGGTGGAAC AGTAGGAGAT ATCGAGTCCT 780

TGCCATTCCT AGAGGCTCTT CGTCAGATGA AGGCAGATGT GGGTGCGGAT AATGTCATGT 840

ATATCCATAC AACCTTGCTT CCTTACCTCA AGGCTGCTGG TGAAATGAAA ACCAAACCAA 900

CCCAACACTC TGTCAAAGAA TTGCGTGGCT TGGGAATCCA ACCAAATATG TTGGTTATTC 960

GTACAGAAGA GCCAGCTGGT CAAGGAATTA AAAATAAACT GGCCCAGTTC TGTGATGTGG 1020

CACCAGAAGC CGTTATCGAA TCGTTGGATG TTGAACACCT TTACCAAATT CCACTGAACT 1080

TGCAGGCACA AGGGATGGAC CAAATTGTTT GTGATCATTT GAAATTAGAC GCACCAGCAG 1140

CGGATATGAC AGAATGGTCA GCCATGGTGG ACAAGGTCAT GAACCTCAAG AAACAAGTTA 1200

AGATTTCCCT TGTTGGTAAG TATGTGGAGT TGCAAGATGC CTATATCTCA GTGGTCGAAG 1260

CCTTGAAACA CTCTGGCTAT GTCAATGATG CAGAAGTTAA AATCAATTGG GTCAATGCCA 1320

ATGATGTGAC AGCAGAGAAT GTAGCAGAAC TCTTGTCTGA TGCGGACGGG ATCATCGTAC 1380

CAGGTGGTTT TGGTCAACGT GGTACAGAAG GGAAAATCCA AGCCATCCGC TATGCGCGTG 1440

AAAATGATGT TCCAATGTTG GGAGTCTGCT TGGGAATGCA GTTGACATGT ATCGAGTTTG 1500

CTCGTCACGT TTTAGGTCTT GAAGGTGCCA ATTCTGCAGA GCTTGCACCA GAAACAAAAT 1560

ACCCTATCAT TGATATCATG CGTGATCAGA TTGATATTGA GGATATGGGT GGAACCcTTC 1620

GTTTGGGACT TTATCCGTCT AAGTTGAAAC GTGGCTCTAA GGCTGCTGCT GCTTATCACA 1680

ATCAAGAAGT GGTGCAACGC CGTCACCGTC ACCGTTATGA GTTTAATAAT GCCTTCCGTG 1740

AGCAGTTTGA GGCAGCAGGT TTTGTCTTTT CAGGAGTTTC TCCAGACAAT CGTTTGGTAG 1800

AAATCGTGGA AATTCCTGAA AATAAATTCT TTGTAGCTTG TCAGTATCAC CCTGAACTGT 1860

CAAGCCGTCC AAACCGACCA GAAGAACTCT ACACTGCCTT TGTTACTGCA GCAGTTGAGA 1920

ACAGCAATTA GCAAAATCAG AACCTTTGAG AAAAATCTCA GAGGTTTTTT GCATACGATG 1980

ATATTGCAGT ATATCTGAGG TAGGGGTCCT CTGTATGTAC CTGCTACCGT TGAAATCAAT 2040

AGCGACTCCC TCTTGCCCTG TGCTAGTGAA TGGATTTATC AGTATATTGA AATGAAATAA 2100

AATTTGAACA AATTAATTCG GAAAGCCAAA TCAATTTCTA GCAAAGTTTT AGGAACTGGA 2160

TTGTATAGTG AATTGAAATA AGATGTGAAC ATCTCTATCA GGAAAGTCAA ATTAATTTAT 2220

AGAAATATTT TAGCAGTCAA GATGTACTGT TATAGATTCA ATACATTATA CTTTTTTAAT 2280

TTAATCCACT ATAGTAAAAT GAAATAATAA CAGGACAAAT CGATCAGGAC AGTCAAATCG 2340

ATTTCTAACA ATGTTTTAGA AATAGAGGTG TACTATTCTA GTTTCAATAT ACTATCCCAA 2400

ATCATTCATA CCTCTCTCAA CTAGATGTAA CTTACAAAAC CCCTGACCTC ATGAGCCACT 2460

TTCTTCCTCC TCATGAGGTC AGTTTTACTT TCTGCTGTTC CAGTATCGTT TTTCCTCGCT 2520

AGATTTCCTC AAAAGGGCAG ACTCCTCCCT TGGTGCGTCA CACGATTTTT TCATCTCGAC 2580

TGTTCTTTAA TGCATCATTA ACGACGCTTT TCTTCTAGGT GGTTCATAAG GAACAGGAAG 2640

ATTCAGGTTG ACTTTTCTAA TCCTAGAATA AAGTGCTGAA AACAATTCGG AATAGGCATA 2700

GAGACTAGAC AATTTGAGGA GCTGCTTGCG TCCTGTTCGA ACACATTTTC CCACCACGTG 2760

AAGA 2764

(2) INFORMATION FOR SEQ ID NO: 241:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1682 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

CCGTTTTTTT CATTGTTCAG TACTACAACT TACGTTGTAG CGCCCTGCAC ATTGGTTCGT 60

CTTGTTCAGT TTTCAAAGGT CTTTGTCACT TGCTTCTCTC AAGCGACAAC TATATTAGTA 120

TATCACAACT GCTTTCGCTT GTCAACACTT TTTTGAAGAT TTTTAAGTTT TTTTAAACTT 180

TTTTTCATCA AGTGGTCCTG ACGCAACATA CCATAGTCCG TACGGGATTC GAACCCGTGT 240

TACCGCCGTG AAAAGGCGGT GTCTTAACCC CTTGACCAAC GGACCTGAGT TGTTATTTTC 300

AACTCTTACT ATTATACAGT CTTTTCAAAC TTTGTCAACT ACTTTTTTAA ACTTTTTTTA 360

TTAATTTTAC AACAGCTTCA GTTCGAGCTG TATGTGGGAA CATATCGACC GACTGGATAT 420

AATGAAGATC ATAGACTTCT ACTAAGCGTA CCAAATCACG AGCCAAGGTC GAAACATTAC 480

AAGAAATATA AACCATTTTT TCTGGTACAT AAGTAAGAAT AGTATCTAAT AACTTATCAT 540

CCAGACCTGT ACGTGGTGGG TCAACAATCA AAGCATCTGC TCGGTAGCCT TCCTTGTACC 600

AACGAGGAAT AATCTCTTCT GCCGTTCCAG CTTCATAATG AGTATTGTCA AATCCCATTC 660

TTTTAGCATT TCGCTTGGCA TCTTCAATAG CTTCTGGAAT AATATCCATA CCTCTGAGTG 720

TTTTTACTTT CTTTGCAAAG GCAAATCCAA TCGTTCCAAC TCCACAATAA GCGTCAATCA 780

AATGGTCTTC TTTATCAACA TCCAGCGCTT TTACTGCTTC GCTATAGAGG ACTTCTGTTT 840

GCTCAGGATT TAGTTGATAA AAAGCTCGAG GGGATAGTGA AAATTCATAA TTGAGTACAC 900

CTTCTTGAAT ACTCTCTTGC CCCCAGATAA TCTCTGTCTT TTCACCATAT ATCTCACTGG 960

TTTTAGCTGT ATTTGTATTA ACAGCTACTG TCACAACTTC TGGGAAATCT TTAACCAACT 1020

CTTTTACCAA TTGAGTTAAA TTAAGCTGGC GGTTTGTAAC AATAATAATC TGAACCTGTC 1080

CGGTCTTTCT CGCGCGTCGG ACCATAATAG TACGGACACC TAGAACTTTT CTCTCATCCG 1140

TGATTGGAAT CTGGTGATAA GTAAGTAATT CTGCTAAGCG ATTAGCAATC ACTTGGGTTT 1200

CCTTATCTTG TACCAGGCAG TCTTTCAACT CTACTAAATA GTGAGAGTTT TGTGCATATA 1260

AGCCCGCCTT GACCTGATTT TTAAATTTTC GAGTCTGAAA TTGTAACTTA GCTCTGTAAT 1320

ATTTTGGTTC CTGCATTCCA ATAGTTGGAC GAATTTCATA ATTTTCATAT CCTGCAGGAG 1380

CAAATTTTTT CAGCGCTTGA TGAAGTAAGT CCGTCTTGAA CTCCAGCTGC TTATCATAAT 1440

GCAGGTGCAT GATTTGGCAG CCTCCGCATT CATTATAAAT AGTACAAGAT GGCACAATTC 1500

GAAATTTAGA CTTCTTGTTG ACCTTCAGTA ATTTTGCTTC AACAAAGTTG CGTCTAATAG 1560

AAGTAATCTG ACAATAGATA TCTTCGCCTT TGAGAGCTCC TGGTACAAAG ACTAATGTTT 1620

TTTGGTAAAA GCCGATTCCC TCACCGTTAA TTCCCATGCG CTTGATTTTT AATGGTATTT 1680

TT 1682

(2) INFORMATION FOR SEQ ID NO: 242:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2524 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

TTAACTTTGG TCAATTCTTT AAAGTCATCC TCTGTAAGCA TGTCTAACCA TTGATGTTTC 60

CCTTTATTGC TAAAATCACC AATTCCGACT ACAGCTATAT CTAAATCTTT CCAACTATTT 120

TTCAAATTTT CAAAATATCT TGATTGCAAA ATACCATCTG CTAACAATTT ATTTTCTTGC 180

ACAATCGTTG CATTCATAAA TGTACACTCT CCATGAAATT TTCTAGACAT TTCATAAATC 240

AGTGTATTCA CATGGTATTT AGCGTGTATG TGACTAGGAC CACCTGCTAG AGGATAGAAG 300

TGAACATTTC GGACACTTTT ACTGTGAATT AAATCTACTA AATTACTTAA ACTTTTCCCC 360

CAAGAAAAGC CAATTTTCAT ATTATCATCA ATTAGATTCC TAAGGACGCC TGCTGCAACT 420

TGAGAAATTC TTTCAGATAA AATTGTTGGA GTATCATCAA ATTCATTTGG AATAATTTCT 480

AAACTTTCCA AACTGTATTT TTCTTTTACA TAATTTTCCA ACTTAAACAT ATTGGTATCA 540

AAATTCTCTA TTTCAATTTT AACAATTCCT ACATTCCTTG CTTCTGTTAA CATTCTACTA 600

ATAGAGGTTC TATAAATTCC TAATTTTGCT GCTATTTGTG ACTGATTTAA GTTTTCAATA 660

TAATACAGAT AAGCAATTTT AGAAAGCAGT TTATTCCTAT CTTGATTCAT ACACTTAACC 720

TCTTACGAAA CTACCTTAAC CATTATCCCA GCATTTTCTA ATGTAGCTAT ATTTTGTTTA 780

GAAAGTTTTT CGTCTGTTAT TACTTCATAG ACTTGACTTA AAGCAAATCT TCTTACTGTA 840

CCTCTTTTAT CAAATTTACT TGAGTCAGTT AGGACAATGA CTTTATCCGA CACTGCTGAA 900

ATATATTGAA CTACCTCACT GCGCATTAAA TCTTTTCCGG TAAAGCCCAT CTCTTTATCG 960

TAACCATCTG TCCCAACAAA AGCTTGACAC ACATGAAAAG TCTGTATCAT TTCTTTTAAT 1020

AAAGGTCCTA CAGTCACCTG TGAATCTTTC TGAAACTCAC CACCAAGAAC AATAACACGA 1080

CATGAATCAT AAGCTCTCAC AAAATTTGCT ATAAAAAACG AATTTGTTAC AATCGTAACA 1140

TTTCTTTTTT GCTTGCAAAT TTCCTCAGCA AGTAAAGCAC AGGTCGATCC AGATTCTATC 1200

ATTATTGTTT CATTATCTGA CACCAATTTT ACTGCTTCCT GAACAATTTT TCTCTTAGTT 1260

TCATAATTAA TTGACAAACG TACATTTAAG TCATCTCCAC TATTTAATAC AGCATATCCA 1320

TGCTCTCTGT GTAATAAACC TTTTGACTCT AATTTATCTA AATCTTTTCT AATCGTTACT 1380

TTCGATACAT TTAATTTTTC CGATAATGTA TTAACGTCGA TCTTTTCATA TTCTGATACT 1440

AATTTAATAA TTTGTTCCAA TCTTTTCATT TTACACCTCC GTTTTATTCT ACCAAAATAA 1500

AAAGCAAAAA ACAACAAATT AACCTTTCGT TCGTAATTGT TTTTCTTTCG TTTTTGTGAT 1560

AGGATAGACT TATGAAGAGG AGGAACTCTT ATGGAAATAT CTAAAGGAAT TATTTTTAAT 1620

ATTCAACACT TTTCAATTCA TGACGGTCCG GGTATTCGTA CAACTGTTTT TTTAAAAGGA 1680

TGTCCTCTGC GCTGTCCATG GTGTTCTAAT CCTGAATCTC AAAGAATGAA ACCTGAAAAA 1740

ATGAAAGATG CTCAACGAGA GAAATTCACC TTAGTCGGTG AAGAAAAGAC TGTAGAAGAA 1800

ATTATTACAG AGGTATTAAA AGACAAAGAA TTTTACGAAG AATCCGGTGG AGGTTTAACT 1860

TTATCAGGAG GTGAAATATT TGCTCAGTTT GAATTTGCTA AAGCCATCTT AAAATCAGCT 1920

AAAGAACATC ACATACACAC TGCCATTGAA ACTACTGCCT TTGTTGATCA TGAAAAATTT 1980

ATTGATTTAA TTCAATATGT GGATTTTATC TACACAGACC TAAAACATTA TAATTCTATA 2040

AAACATAAAA AAGTGACTGG GGTTTTTAAT CAAATGATTA TTAAAAACAT TCATTATGCT 2100

TTTTCACAAA ATAAAACTAT CGTTTTAAGA ATCCCAGTTA TTCCTAATTT TAACAATAGT 2160

TTAGAGGATG CAGAAAAATT CGCTACTCTA TTTAACTCAT TAAATATCGA CCAAGTTCAA 2220

CTACTCCCTT TTCATCAATT TGGTGAAAAC AAATATCGTT TATTAAATCG GAAATATGAA 2280

ATGGATGGAA TCAACGCACT TCATCCwGAA GATCTTATTG ATTATCAAAA GGTATTTCTG 2340

AACCACCATA TTAATTGTTA TTTCTAGTTT ATTTCCTTGA AATGCTCTAG CTATTTGCAG 2400

ATAACAAGCA TCTATAATAC ATACTTAACT TTTCAAAAGG TTTAGCTAAA AAATTTTAGC 2460

CAAACCTTTT CTATTTTACC TTGCTCTAGA ATTTTTAAAC TGCTATACTT ATCACAAAAA 2520

AACG 2524

(2) INFORMATION FOR SEQ ID NO: 243:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2359 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

CGTGCTTGGG GGCTTGTGGT CAAAAGGAAA GTCAGACAGG AAAGGGGATG AAAATTGTGA 60

CCAGTTTTTA TCCTATCTAC GCTATGGTTA AGGAAGTATC TGGTGACTTG AATGATGTTC 120

GGATGATTCA GTCAAGTAGT GGTATTCACT CCTTTGAACC TTCGGCAAAT GATATCGCAG 180

CCATCTATGA TGCAGATGTC TTTGTTTACC ATTCTCATAC ACTCGAATCT TGGGCAGGAA 240

GTCTGGATCC AAATCTAAAA AAATCCAAAG TGAAGGTCTT AGAGGCTTCT GAGGGAATGA 300

CCTTGGAACG TGTCCCTGGA CTAGAGGATG TGGAAGCAGG GGATGGAGTT GATGAAAAAA 360

CGCTCTATGA CCCTCACACA TGGCTAGATC CTGAAAAAGC TGGAGAAGAA GCCCAAATTA 420

TCGCTGATAA ACTTTCAGAG GTGGATAGTG AGCATAAAGA GACTTATCAA AAAAATGCGC 480

AAGCCTTTAT CAAAAAAGCT CAGGAATTGA CTAAGAAATT CCAACCAAAA TTTGAAAAAG 540

CGACTCAGAA AACATTTGTA ACACAACATA CAGCCTTTTC TTATCTAGCG AAGAGATTTG 600

GGCTTAATCA ACTTGGTATT GCAGGTATCT CTCCTGAACA AGAACCAAGT CCACGACAAC 660

TAACAGAAAT TCAGGAATTT GTTAAGACCT ATAAGGTTAA AACGATTTTT ACAGAAAGTA 720

ACGCTTCTTC AAAAGTAGCT GAAACTCTTG TCAAATCAAC AGGTGTGGGT CTTAAAACTC 780

TGAATCCTTT AGAGTCAGAC CCACAAAATG ACAAGACCTA TTTAGAAAAT CTTGAAGAAA 840

ATATGAGTAT TCTAGCAGAA GAATTAAAGT GAGGAAAGAA TGAAAATTAA TAAAAAATAT 900

CTAGCAGGTT CAGTGGCAGT CCTTGCCCTA AGTGTTTGTT CCTATGAGCT TGGACGTTAC 960

CAAGCTGGTC AGGATAAGAA AGAGTCTAAT CGAGTTGCTT ATATAGATGG TGATCAGGCT 1020

GGTCAAAAGG CAGAAAACTT GACACCAGAT GAAGTCAGTA AGAGGGAGGG GATCAACGCC 1080

GAACAAATTG TTATCAAGAT TACGGATCAA GGTTATGTGA CCTCTCATGG AGACCATTAT 1140

CATTACTATA ATGGCAAGGT TCCTTATGAT GCCATCATCA GTGAAGAGCT CCTCATGAAA 1200

GATCCGAATT ATCAGTTGAA GGATTCAGAC ATTGTCAATG AAATCAAGGG TGGTTATGTC 1260

ATTAAGGTAA ACGGTAAATA CTATGTTTAC CTTAAGGATG CAGCTCATGC GGATAATATT 1320

CGGACAAAAG AAGAGATTAA ACGTCAGAAG CAGGAACGCA GTCATAATCA TAACTCAAGA 1380

GCAGATAATG CTGTTGCTGC AGCCAGAGCC CAAGGACGTT ATACAACGGA TGATGGGTAT 1440

ATCTTCAATG CATCTGATAT CATTGAGGAC ACGGGTGATG CTTATATCGT TCCTCACGGC 1500

GACCATTACC ATTACATTCC TAAGAATGAG TTATCAGCTA GCGAGTTAGC TGCTGCAGAA 1560

GCCTATTGGA ATGGGAAGCA GGGATCTCGT CCTTCTTCAA GTTCTAGTTA TAATGCAAAT 1620

CCAGCTCAAC CAAGATTGTC AGAGAACCAC AATCTGACTG TCACTCCAAC TTATCATCAA 1680

AATCAAGGGG AAAACATTTC AAGCCTTTTA CGTGAATTGT ATGCTAAACC CTTATCAGAA 1740

CGCCATGTGG AATCTGATGG CCTTATTTTC GACCCAGCGC AAATCACAAG TCGAACCGCC 1800

AGAGGTGTAG CTGTCCCTCA TGGTAACCAT TACCACTTTA TCCCTTATGA ACAAATGTCT 1860

GAATTGGAAA AACGAATTGC TCGTATTATT CCCCTTCGTT ATCGTTCAAA CCATTGGGTA 1920

CCAGATTCAA GACCAGAAGA ACCAAGTCCA CAACCGACTC CAGAACCTAG TCCAAGTCCG 1980

CAACCAGCTC CAAGCAATCC AATTGATGAG AAATTGGTCA AAGAAGCTGT TCGAAAAGTA 2040

GGCGATGGTT ATGTCTTTGA GGAGAATGGA GTTTCTCGTT ATATCCCAGC CAAGGATCTT 2100

TCAGCAGAAA CAGCAGCAGG CATTGATAGC AAACTGGCCA AGCAGGAAAG TTTATCTCAT 2160

AAGCTAGGAA CTAAGAAAAC TGACCTCCCA TCTAGTGATC GAGAATTTTA CAATAAGGCT 2220

TATGACTTAC TAGCAAGAAT TCACCAAGAT TTACTTGATA ATAAAGGTCG ACAAGTTGAT 2280

TTTGAGGCTT TGGATAACCT GTTGGAACGA CTCAAGGATG TCTCAAGTGA TAAAGTCAAG 2340

TTAGTGGAAG ATATTCTTG 2359

(2) INFORMATION FOR SEQ ID NO: 244:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1052 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

TTCTTTCTGC TATAATCGTA TAAAATACTT ACTTTAGGAG TTCTTATGAA AGTTGTTAAA 60

TTTGGAGGTA GTTCTCTTGC CTCTGCTAGT CAATTAGAAA AAGTTTTAAA CATCGTCAAA 120

AGCGATTCAG AGCGTCGTTT TGTAGTCGTT TCTGCGCCTG GTAAACGCAA TGCTGAAGAT 180

ACTAAGGTTA CGGATGCCCT GATTAAATAC TACCGCGACT ATGTTGCGGG TAACGATATT 240

AGCAAGAACC AAAGCTGGAT TATCGACCGC TATGCTGCTA TGGTTAGTGA ATTGGGACTA 300

AAACCAGCTG TGCTAGAAAA AATTTCTAAA AGCATTCACG CCTTGGCCAC TCTTCCTATT 360

GAAGAAAATG AATTTCTCTA CGATACTTTC CTAGCAGCCG GTGAAAATAA CAATGCCAAA 420

TTGATTGCTG CCTACTTTAA CCAAAATGGT ATCGATGCAC GCTATATGCA CCCTAGAGAA 480

GCTGGGATTG TGGTCACAAG TGAACCTGGT CACGCTCGCA TCATTCCATC AAGTTATGAC 540

AAGATTGAAG AATTGACAAA CACCAATGAA GTCCTTGTCA TTCCTGGTTT CTTTGGTGTC 600

ACTAAGGAAA ATCAAATCTG TACTTTCTCA CGTGGAGGTT CTGATATTAC AGGTTCTATC 660

ATTGCTGCTG GTGTCAAAGC TGACCTCTAT GAAAACTTTA CGGACGTTGA TGGTATCTTT 720

GCAGCCCACC CTGGTATTAT CCACCAACCA CACTCGATTC CTGAGTTGAC CTACCGTGAA 780

ATGCGCGAGT TGGCCTATGC AGGCTTCTCA GTCCTTCATG ACGAGGCTCT TCTTCCTGCC 840

TACCGTGGAA AAATTCCTCT GGTTATCAAG AATACCAACA ACCCTGACCA TCCAGGTACT 900

CGTATCGTTC TAAAACACAG TAATGATGAA TTTCCAGTTG TGGGAATTGC TGGTGACTCA 960

GGCTTTGTCA GCATTAACAT GTCGAAATAC CTCATGAACC GTGAGGTTGG ATTTGGCCGC 1020

AAGGTTCTGC AAATCCTGGA AGAACTTAAC AT 1052

(2) INFORMATION FOR SEQ ID NO: 245:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 855 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

CCCTCGAAAA CTAAGCCGAT GAAGTCAGAA CACTTCAATC CTGTTCGTGA CTGGTGGGAA 60

AATCGTGAAG AGATTCTGGA AGGTAAGTTC TACAAATCTA AATCATTTAC ACCTAGTGAA 120

TTGGCTGAGT TGAATTATAA TTTAGACCAG TGTGACTTTC CAAAAGAGGA AGAGGAAATC 180

TTAAATCCCT TTGAGTTGAT TCAGAATTAT CAAGCGGAAA GAGCAACTTT AAATCATAAG 240

ATTGATAATG TATTAGCTGA TATTTTGCAG TTGTTGGAGG ACAAATAATG ACACCAGAAC 300

AACTTAAAGC AAGTATTCTC CAAAGAGCGA TGGAAGGGAA ATTAGTGCCG CAAAATCCCA 360

ATGACGAACC TGCAAGTGAA TTATTAAAGA GAATTAAAGC TGAAAAAGAA AAACTTATCA 420

GTGAAGGAAA AATCAAACGA GATAAAAAGG AAACTGAGAT ATTTCGTGGT GATGATGGGA 480

AACATTATGG GAAGTTTGCT GATGGAAGCA CTCAAGAAAT TGATGTTCCT TATGATATTC 540

CTGATACTTG GGAGTGGGTG AGGATAAAAT CAATTTATTG GAATTTTGGG CAAAATAAGC 600

CAGAGAAATC CTTTAGGTAT ATAGATACGT CTAGTATTGA TAGAAAAAAG AACATAATCA 660

ACTACAAAAA TCTACAATAT CTTTCACCTG AACAAGCGCC TTCCCGTGCT AGAAAATTAG 720

TTTCGCAGAA TAGTGTCTTA TTTTCAACAG TTAGACCATA TCTAAAAAAT ATTGCTGTAG 780

TTAGAGAACT TAAAGAGTAT TTGATAGCTA GTACAGCATT TAATGTTTTG GGATACTTTA 840

CTTAACGAAA CATAT 855

(2) INFORMATION FOR SEQ ID NO: 246:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 660 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

TTTAGGAAGG CTATCCGTAA TTTTACAAAG GATTTAGATA TTACAGAGGA ACATTTAGAT 60

ATTATCAAAA GAGAGATGTT TGGCGAATTT TTCAGTAGCA TGAACTCTCT TGAATTTATT 120

GCAACGCAAT ATGATGCTTT TGAAAATGGT GAGATAATTT TTGATTTGCC GAAAATTTTA 180

CAGGAAATTA CTTTAGAGGA TGTCCTTGAT GCTGGACATC ATTTAATAGA TGATGGTGAC 240

ATAGTTGATT TTACAATATT CCCATCGTAG TAACCTATTA TAATAGACAC TAGAAAGAAG 300

GGATGACAAG TATGAGAAAA AAAACAATTG GAGAGGTTTT ACGATTAGCT AGAATCAATC 360

AGGGATTGAG TTTAGATGAA TTGCAGAAAA AGACAGAAAT CCAGTTAGAT ATGTTGGAAG 420

CAATGGAAGC AGACGATTTC GATCAACTTC CAAGTCCTTT TTACACGCGT TCTTTCTTGA 480

AAAAATATGC ATGGGCTGTT GAGTTAGATG ACCAAATTGT TTTGGATGCT TATGATTCTG 540

GGAGTATGAT TACTTATGAG GAAGTAGATG TTGATGAAGA TGAGTTGACA GGTCGTAGAC 600

GTTCAAGTAA GAAAAAGAAG AAAAAAACAT CATTTTTACC TTTATTTTAT TTTATCCTGG 660

(2) INFORMATION FOR SEQ ID NO: 247:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1805 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

CCGGTTGCAC AGGATCGTGC ATAGTCAACT CTTCAAGTAT AGCATATCTC CTATTTTCTT 60

ACAAGTAATA ACACCTAAAA TGAAGCTTTT TCTTTTACTT TTTTCTGCCA AGAGGCAAAA 120

AGCATGCTGA GGTAAAAAAC GCTCATCATA ATAGGAACAC CAAGAATGGT CTTTTCATGA 180

TAGAAAATCG TCAAATAGGC TGAAAAGACA ACGCCAAGGA CAAAACTACT AAGCAGGCTA 240

ACAAATATGA ATCCTTCACG CAAAAAAGGA GTGTGCTTGG TTCGGAAATA ATCTCCAAAA 300

GCCAGCATGG TCCGTTTGAT ATTCCCTGTC ATAAAAGCGT TATTATAGGC AATACCCGAC 360

ACTTCTCCAA AAGCAGTTGT CACCAGTCCC ATACAGAAGG CCAAGGGCGG CACTAGATAG 420

ATATTATCCA CAGTTTGCGG CACAAAAGCA ATAATGATTG ATAAGATTGC CAAGGGAATC 480

AAGGACAGAA TAGGTTTTTT CACAATTCTC AATTTTTCCT TATAAATCGT TAATAAAAAG 540

ACTCCCATCA TAAACGCTAG CAAGGTGAGA ACCTTGTCCC TAACATCCGA AACATTATTT 600

TTAATTAATT CTACTGAAAG AAAGACAACA TTTCCAGTTT GTCCAGCTAC AAGGGTATTC 660

CCGCGAACAA TAAAAGTGTA AGCATCCACA TATCCAGCAC AAAACGTCAA AAAAAGTGCT 720

AACCTTTTAG ACTGACGTGA TATTTTTCTT ATAGGTAATA ACCTCATTTT ACCTCCCATT 780

GTATTTTCTC TTAGAAATAT TGTACCATTT TCTTTCTAAA AAATCGTAGG CTACCATTTA 840

GATTTTACTA TTAGCATAAA AATAATAATA GACAACTATT TATCCAAAAA TAGATAGATG 900

TAACATGTTT GCAAACAAAG CATACGAACC TTTAGTAAAA TCATTTCCAT GAAACTAGAA 960

TAGAGCCCTC TTAGCAAAAA TCATTATTTT AATTTATTTC TAATCACTCC TTGACATAAA 1020

TAACTCTCAC CAATAAAAGA CTATGTCTTA AAAAAATGGT ATAATAAAAT CAATACTTGG 1080

GCTTGATGGC TATGCTACTA ATAACAATTA GGAGAGAAAA TCAGGCACTT GTTAACAACA 1140

AGGATTATCC CCTTGAGATG AAAGGAACTT TAGAAATCTT ATGATGAACA TGCAAAACAT 1200

GATGCGTCAA GCACAAAAAC TTCAAAAACA AATGGAACAA AGCCAAGCTG AACTTGCTGC 1260

TATGCAATTT GTTGGCAAAT CTGCTCAAGA TCTTGTCCAA GCGACCTTAA CTGGCGATAA 1320

GAAAGTTGTC AGCATTGATT TCAATCCAGC TGTCGTTGAC CCAGAGGACC TTGAGACTCT 1380

TTCTGATATG ACCGTTCAAG CCATCAACTC TGCTCTTGAA CAAATCGATG AAACTACCAA 1440

GAAAAAACTG GGTGCTTTCG CTGGGAAATT ACCTTTCTAA AAACAAGGAG CTAGAACAAT 1500

GCTTGTCGAT AACAAAGGCT AAGAAAGGTG CAAAAATGAC TCTATAATAT TTGTAGTGGG 1560

TAAATCCCCT ATGGATATTA TGGAGCCTAT TTTTGTGTAG AAAAAAGTCC CATATGACCT 1620

ATAATGAAAA GCGACAAAAC AACTCATTAG AAAGAATCAT ATGGAACAAT TACATTTTAT 1680

CACAAAATTA CTAGACATTA AAGACCCTAA TATCCAGATT TTAGACATCG TCAATAAGGA 1740

1270

TACACACAAG GwAATCATCG CCAAACTGGr CTATGAAGCT CCATCTTGTC CTGAGTGCGG 1800

AAGTC 1805

(2) INFORMATION FOR SEQ ID NO: 248:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2516 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

CTGCATCTAG TTTGTTTCTC CCTACAGTTT TAGCTAGACA GATTGGAGAT TATGATTTAA 60

CGTCGCCGCG TTGGGGTTCG GATACAACTA GTGAGCTTGA GAAAGAAAAC TCCTCTGCTG 120

GAATTAATAA TAATGACAGC ACTGGTGGCG GTAAAAGGTT AAATACCTCT ATTCGTAGCG 180

CCTATAGTGG GTCAGATATT ACCCCGGTAT ATTCATTGGG GTCTGGCTCT AGGATTGTCA 240

TGTACTATAA TGGAGGTGGT GACAATTATA TTGGTTCTGG TACTAGATTA GCTATGGCGC 300

CACAATTTGG AAATCATGTA AGAATTCATA CTTCAGGTTC TTGGAATCCA GATTCTTATT 360

AACTTACTTG TCAGAGTAAG CCTTAAAGAT GGTTGATTGT GGGTGTAGCA TGAAAAAAGA 420

ATGCTACACC CTATTTTTAT TATAAGGAGG AGTAAGGATG GAATTTTTCA TTTGTAATCT 480

TGTACGAGTC GTTCAATCAC CTCGATTTTA TATGTCTTTA TTTTTGACCC TTCTTTGCAT 540

GAGTTTAGGA AATTTCCTTG CTTTCAATGG TATTTATAAA ATTGAAGGTT TATCGATTTT 600

TTTTGCCGCT TCTTCTATTC GAGGATTTTC ACCGATTAGC CTAGTAGCTG CACTTATCTG 660

TACACTGCCC TATTCTAGTC AGATAATAGA GGATGCTGAG AGTCATTTTC TAACAGCACA 720

ATTGTGTCGA ATTTCTAAAA AGAAGTATCT GGCTATTGTG GGTAGTACTG TAATTATTTC 780

TTCTTTTCTA GTCTTTTTTC TCCCCTATTT ATTATTATTA GGAATTAATC TTTTAGTGAC 840

TCCTTATCAG GAAATTTATA TTGGAGATTA TAGTGGTGCC TTAAAAGAAT TATTTGATTC 900

CAATCAGTTT CTCTATAGTC TTGTAACGAC TCTCTGGTAT GGAGTTTGGG GCGCTGTGTT 960

CTCTATTTTT GGACTAGCTA GTGCTTTGCT AGTGAAGAAA AAAATAGGAG CTATTTTCAT 1020

CCCAGTTGCC TATATGATGG TTGGTGGTAT TTTTTGGGCT ATTTTAGGGC TATCTTACTT 1080

AGAACCTGTG ACAACGCTAG CTTTGGGATA TCAGAAAGAT ATCAGTCTTT CCTTAGTTAG 1140

TGCTCATCTT GCTTTTATTT TATTTGTTAG TTGTTTGGTT GTTTATGGTA CATTTTTTCT 1200

ACATTCAGAG GACTATGTAT AATGAAACAA TTTGTTCAAT TTTATAAAAA AGATTTCTTA 1260

GCAGTATTGG TTTATTTTAT ATTACTGCTA TCCTGTGTTT TATCTAGTAC AGTATATTTA 1320

TTGCGCtGTC GCCAATATTC AATCCATCCA AATGTATTAG AATGGATCTT AGTTTTACTT 1380

CAAGATATGA CGACTGGAGT ATATTGCTTT CCGTTCACAT ATATATTGTT CTTTTTTTAT 1440

TTGATGAATA ACTATTTTAA TAGGTTGGAG TGTCGCATTC GTCTGAAATC AATTAAGCAC 1500

TTTACCAGTT TTAGTTTCAA ATTAGCAGCT CTTAGTACGG GGATTTGGAC GGCGACTTTA 1560

TTTTTATTGA TTTTTCTAAT TGCATTTAGT AATGGTTTTA GCTTCTCTTT GGAGATAAAG 1620

GAGGTTGATT TTTTAAGAGA ATTTTATGGT ATAAGTATTG CAAACAATGC TAGTTTCTTT 1680

ATAGGATTTT TTTTCTCTTA TATAGCATAC TATTTCTTTT TATCCTTACT TACTATTAGC 1740

AGTTTTTCTT GGTTTAAAAA ATCAAACATG AGCTTAGTAT TTCTGTTTAC TTTTTTATTT 1800

GTAGAATCCT TATTCTGGAT TTATCAGTTG GACAATGGGA TAATTGGATT ATTGCCAATT 1860

TTTCAGTATA TGGTAAATTC CAATCCGTAT GCATTGATTT ATTGGCTTAC ATTACTATCT 1920

ATCATAATTC CATTGACTGT ATTTTCTGTT CATAGAAACT GGAGGAGAGT GTAAAAGTTG 1980

GAAATGGGAA AGTTAAGTAG TCACATGTGG AGGTTGAATC AGATAATCTA TACCAAGTAC 2040

TTTTGGGGTT ATGTTCTTTT TTGGATATTG ATTTGTTTAG GATTATGGTA TTGGTTAGAA 2100

GGAAATGATA GACTTGTTAT AGAAATTTTA AAAGGGCCTA ATCTGAGTCA AAACTCTTTT 2160

TTAGTCTTAT CTATATGGTT GCTTCATTGG TTTATTATTC ATACATTTTT TCTAGCAGTT 2220

GTATATCGTA GAAGAGCATC CGATTTCTTT ATGGAAGTGA TTCGATTTTC TTCTATTAAG 2280

CTCTGGATTA GGTATCAGAT TTGGACCTGT TTTCTTTATG GACTCATTTT AATCATGGTA 2340

AAAGTTCTAG TGATTCAATT TATGTTACAG TTACCAAACT GGGATATAGG AGTTTTGTTT 2400

ATAGTTGATT CTTTGAATGC TTGTGTGTTA GTCTTGTTTT GCTTTATGTT ATACGCACTA 2460

GGAGCGAATG TACAAATGAA CTTTGCTTGC GTTAGTTTCT TTTTACTCAT GATTGG 2516

(2) INFORMATION FOR SEQ ID NO: 249:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1364 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

CGGTGTTTTT TTGTAAATTT TCTAGCACTT GTATGGTAAA ATAGATACAG GTGTTCATTA 60

AACTAGACTA AAAACCTATT TAAGCAGGCA AAATGAAGAA ATACCAACAA TTATTTAAGC 120

AAATCCAAGA AACCATTCAA AACGAGACTT ACGCTGTCGG AGATTTCCTT CCTAGCGAGC 180

ACGACCTTAT GGAGCAATAT CAAGTGAGTC GTGATACCGT CCGAAAGCCC TGTCTCTCCT 240

CCAAGAGGAA GGATTGATCA AAAAGATAAG AGGGCAAGGT TCTCAAGTCG TCAAAGAAGA 300

AACCGTCAAT TTCCCTGTAT CCAACCTAAC CAGCTACCAA GAACTAGTTA AAGAACTTGG 360

ACTGCGCTCT AAAACCAACG TGGTCAGTCT GGACAAGATT ATTATTGATA AAAAATCCTC 420

ACTGATAACC GGTTTCCCAG AGTTTCGGAT GGTTTGGAAG GTGGTCCGCC AGCGTGTGGT 480

GGATGATCTG GTATCCGTTC TGGATACGGA CTATCTGGAT ATGGAACTCA TCCCAAATCT 540

CACTCGCCAA ATTGCTGAGC AGTCTATCTA TTCTTATATA GAAAATGGCC TCAAACTCCT 600

TATTGATTAT GCTCAGAAGG AAATCACCAT TGACCACTCA AGCGACCGAG ACAAGATTCT 660

CATGGACATT GGCAAAGACC CTTATGTCGT TTCGATTAAA TCAAAAGTCT ATCTCCAAGA 720

CGGACGCCAA TTTCAGTTTA CCGAAAGTCG CCATAAGTTA GAGAAATTTA GATTTGTAGA 780

TTTTGCAAAA CGCAAGAAAT AAAAGACTGA GACACCAGAT CTCAGCCTTT TTCGGCTCTA 840

TAATATTTGT AGTGGGTAAC CCCCCTATGG ATATTATGGA GCCTATTTTG TGTAGAAAAA 900

AAGTCCCATA TGACCTATAA TGAAAAGCGA CAAAACAACT CATTAGAAAG ATTCATATGG 960

AACAATTACA TTTTATCACA AAACTGCTCG ATATTAAAGA CCCAAACATC AAGATTCTAG 1020

ACATCATCAA TATGGATACC CACAAAGAAA TTATCGCTAA GCTGGATTAT GAGGCTCCAT 1080

CTTGCCCTGA TTGTGGAAGT CTAATGAAGA AATATGACTT TCAAAAACCG TCTAAGATCC 1140

CTTACCTCGA AACAACTGGT ATGCCTACTA GAATTCTCCT TAGAAAGCGT CGTTTCAAGT 1200

GCTATCATTG TTCTAAAATG ATGGTCGCTG AAACTTCTAT CGTCAAGAAG AATCATCAAA 1260

TTCCTCGTAT TATCAACCAA AAAATTGCGC AAAAGTTGAT TGAGAAGATT TCTATGACCG 1320

ATATTGCTCA TCAGCTGGCC ATTTCAACTT CAACTGTCAT TCGG 1364

(2) INFORMATION FOR SEQ ID NO: 250:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1227 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

CCATGAAGAC CGCTTGGAAT TGGAATGGCA CAAGTCTTTG TTGAATGGTC TATTCCCATT 60

GACAATCGGT GGAGGAATTG GACAATCTCG TATGGCCATG TTCCTACTTC GCAAGAGACA 120

CATCGGAGAA GTGCAAACAA GTGTTTGGCC TCAAGAAGTC CGCGATACTT ACGAAAATAT 180

TTTGTAGAGA ATCGAACCGC AAGGTTCGGT TTTCTTTCTC TTTTTGTCTA TAATTTGGTA 240

TAATAAACAG TATGAAAATC GTATCAGGAA TCTATGGGGG ACGTCCCCTC AAGACACTAG 300

AAGGCAAGAC GACAAGACCT ACTTCGGATA AGGTTAGGGG AGCCATTTTT AACATGATTG 360

GTCCCTACTT TGAAGTGGGA CGAGTCTTGG ACCTTTATGC AGGTAGTGGT GGTTTATCTA 420

TCGAAGCAGT ATCGCGTGGC ATGTCCAGTG CTGTTTTGGT GGAGCGAGAC CGTAAGCTCA 480

GACCATCGTG GCTGAAAATA TCCAGATGAC CAAGGAAGTT GGAAAATTTC AACTCCTCAA 540

GATGGATGCA GAAAGGGCAT TGGAACAGGT ATCTGGGGAA TTTGACCTCG TTTTCTTAGA 600

CCCTCCCTAT GCCAAGGAAC AAATCGTAGC AGATATTGAA AAAATGGCTG AGAGAGAGCT 660

TTTTTCTGAA GATGTTATGG TTGTGTGCGA GACGGATAAA GCCGTTGAAC TTCCAGAAGA 720

AATTGCCTGT CTGGGTATCT GGAAGGAAAA GATTTATGGA ATTAGTAAGG TGACAGTCTA 780

TGTCAGATAA GATTGGCTTA TTCACAGGCT CATTTGATCC GATGACAAAT GGGCATCTGG 840

ATATCATTGA ACGGGCGAGC AGACTTTTTG ATAAGCTTTA TGTGGGTATT TTTTTTAATC 900

CCCACAAACA AGGATTTCTC CCTCTTGAAA ATCGTAAACG GGGGTTAGAA AAGGCTGTGA 960

AACATTTGGG AAATGTTAAA GTCGTGTCTT CTCATGATAA ATTGGTGGTC GATGTCGCAA 1020

AAAGACTGGG GGCTACTTGC CTAGTGCGAG GTTTGAGAAA TGCGTCGGAT TTGCAATATG 1080

AAGCCAGTTT TGATTACTAC AATCATCAGC TGTCTTCTGA TATAGAGACT ATTTATTTAC 1140

ATAGTCGACC TGAACATCTC TATATCAGTT CATCAGGCGT TAGAGAGCTT TTGAAGTTTG 1200

GTCAGGATAT TGCCTGCTAT GTTCCCG 1227

(2) INFORMATION FOR SEQ ID NO: 251:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3652 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

CCGGTCAAGT TAAAAACGCT ATTTCTTCCC ATTTTATTTA TTTTTTAGGA GTGGTAACGT 60

ATCAAAATAG CCCAAGCGTT CTCACCCGTG TGAGTTTGAA TAATGGAACC CGTTTCCAAA 120

ACAGAAATTG GCTTTTCAAC ATAAGCTTGT AAGCTTTCTT TCATCTCTTT TGCCCAATCA 180

TCACTACCAG AATATGAAAT TCCAATCTCT GCTACAGCAC GTTCAGAAAG CGATGTTATC 240

AACTCATCTA ACCATTTTTT AAATGTTTTA GTTCCACGAC CTTTAACCAT TGGCTGCAAT 300

TCATGGTCTT TCATTTGCAT GACAGCACGG ATATTGAGAA GAGAGCTCAA CAAGCCAGTT 360

ACACGGCTAA TTCGTCCACC TTTGACAAGA TTTTCCAAAG TTGAAACACC AATATAAAGC 420

TCTGTATGGT TTTTAACCTC TTCTACATGA GATAAAATTG CCTCCATATC TTTACCTTCT 480

TGAGCTAACT TCGCAGCCTC AACAACTTGG AATTTCAGGG CTTGGTCAGT GAAGGAACTA 540

TCAACAACAG TCACATCTGC AGTAGATAGG CTAGCACCTT GGCGTGCTGC TTCTACCGTA 600

CCCGAAAGAG CATGGGACAT ATGAATAGCA AGAATCTGGC CACCATCTTT GCATAGGTCT 660

TCAAAAATCT CAGCAAAGAC ACCTACAGGT GGCTGACTTG TTTTCGGAAG ATTCTTACTT 720

TCTTGCATCA ACTGAAGAAA TTTACCTTCT TCTTTCAAAT CCGCATCAGA ATAAACAACA 780

TTATCAATCA TTACAGATAA TGGAACAATT GTAATATCTA ATTGCTTTAC TAGTTCAGGT 840

TCAATAGTAA CAGATGAATC GGTTACAATC TTAATTTTTG TCATAGTATC AATCTTTCTA 900

TTTTAGGATT CAGATTGGTT TCCTTACTTC TAATTATATC AAAAAAAAGA TTAAAAATCC 960

TAATGGAGTC AATCAAATTT TCCGTAAAAT TTGATATAAT CAACTTATAA GAAAAGAGGT 1020

GTCCTATGAT TAAAAAAATT TACCCCATTT TTACCATTTT ACTAGGTGCT GCTATTTATG 1080

CTTTTGGACT GACTTATTTT GTAGTTCCCC ATCATCTCTT TGAAGGAGGG GCGACAGGCA 1140

TTACCCTCAT CACCTTTTAT CTTTTTAAAA TCCCTGTTTC CCTCATGAAC CTGCTGATTA 1200

ATATTCCCCT TTTCATCCTA GCTTGGAAGA TTTTTGGAGC CAAATCCCTC TATTCTAGTT 1260

TACTAGGAAC CTTAGCTTTG TCCGGCTGGT TAGCTTTTTT TGAGCATATT CCCCTTCATA 1320

TTGATCTTCA AGGTGATTTA CTAATCACAG CCCTTATAGC GGGAATCCTA TTGGGAATTG 1380

GCCTTGGAAT TATTTTTAAT GCTGGAGGTA CAACTGGCGG AACTGATATT CTAGCTCGTA 1440

TTCTCAACAA ATACACTCAT ATATCCATAG GAAAACTGCT CTTTATCTTA GATTTTTGTA 1500

TTCTCATGTT GATTCTCCTA ATCTTCAAGG ATTTGAGATT GGTTTCCTAC ACGCTTTTGT 1560

TTGATTTTAT TGTTTCTCGT GTTATTGATT TGATTGGTGA AGGAGGATAT GCCGGCAAAG 1620

GCTTTATGAT TATCACAAAA CGTCCTGACC AACTTGCTAA GGCGATTAAT GATGACCTCG 1680

GAAGAGGTGT TACTTTTATT TCTGGTCAAG GCTACTATAG TAAAGAAAAT TTGAAAATCA 1740

TCTACTGTAT TGTCGGAAGA AATGAAATTG TGAAAACGAA GGAAATGATT CATCGAATCG 1800

ATCCTCAAGC CTTTATAACT ATTACAGAAG CCCATGAAAT CCTAGGAGAA GGCTTCACCT 1860

TTGAAAAAGA ATAAAAAGAG GTAATGTCGT GACCTCAAAA GTTAGACTAA ATCATCTATC 1920

TTTTGGGTTA CAGACAACCT CTTTTTTATT TTATTTACTC AAGCTCTTAA GACCAATTCC 1980

GAGTTACTTC TTCATCAGCC TTTAACTGAT CCACTAATTG GTCAACTGAG TCAAATTTGG 2040

TCATATCTCG AATGCGATCA AGCCAATAAA CCATGACGGT TTCCCCATAA ATATCTTGAT 2100

TAAAATCAAA AATATTGACT TCAAAACGTG CTTCTTCTCC ATCAAAGGTC ACATTTTTCC 2160

CGACACTAGC CATAGCACGA TACTTCTGTC TTTGAATCTC AACATCAACA ACATAAACGC 2220

CATCTGCTGG CATATAAGTA CGGTCTAAAA GCACTAAATT CGCTGTCGGA TAACCAATTG 2280

TACGACCACG AGCATTACCA TGAACCACCA TACCTCTTGA TGGAAGCGGT GCCCCCAAAA 2340

GTTTTCCTGC TTCTTTCACA TTTCCATCTA AAATAGCTTG ACGGATACGA GTTGAACTAA 2400

TCTTTCCTTT CTCATCTTCT ACAGGTGGAA CAATGATAAC TTCTCCATCA AAGTAATTCT 2460

TTAAATCTTC TGCTGTTTTT TTGTCAGAAC CAAATGTATA ATCAAAACCT GCAACAATAA 2520

TTTTGGCATT CATAGCCTTG ATATAAGTTG CAAAGAATTC TTGTGCAGTG AGACTAGCGA 2580

ATTGACTACT AAAATCAAGG AGATATAATT CTTCTACACC TTCGCGCTTT AATTTTCTTT 2640

CACGTTCAGC AGGGTTCAAA ATATGCAAAA ACAAATCTGG ATGATAAGGC TCTAAAGCGA 2700

TCTTTGGAGA TTCATTAAAG GTCATAACGA CGATAGGCAA CAAATCCTTT CTCGCAGCCT 2760

TGTTGGCAAC ACGAAATAAT TCTTGATGCC CCTTATGTAT GCCATCAAAA TAGCCGAGAA 2820

CAACGACTGA ATCAGATGGT GTGCCAATAT CTTTTTGGTT TTTTATAGGA ATAGTAATAA 2880

TCATAAAATA ATTATATCAT AGCGATAGCT ATTTCTGGAA CAGAAAATCT GAAATGTTGT 2940

TTTTTTCACA TGAAGTGTAC CTGTTTTCAA AAAGCACTTT ATTCTATCGT TGCTTAACTA 3000

TGAACTTTGC AATATTCTTC TCAAAAACTT GTAGGACATC TTCAAAATTT TGCAAGGAGT 3060

GATTAGACTT GTTCGGTAAC CATAAAGTGT CATACTATGC TTATGTATGA AAAAGCAATG 3120

CAACTAACTC CTGAGAACTT TAAATTACTA ATTGGTGCCG AAAAGGTAGA ATTTAGAATC 3180

GAGGTACACC TATGGCTGTA AAATTTACAA AATGAGACAA CTTGGGCAAG ATGTTTGAAG 3240

AATTTCCTAA ACTCCCTGAT TTGAAGCAAG TCACTTTCCC TAATGACAAA GAAAAAAGCC 3300

AAAACAGCAA AGAAAAACTA GATGACTGCT TTCCAACAAC TCCCATCTAG TGTGCTTCAG 3360

ACTGGGCTAT TTTTCTCTCC ATCTGTTAGC TTGGATTCTC AGACCGTTTC AGCTAAAGAA 3420

TATCTTTTCC CTTATCAGAA GGAACGGCTC AAGCCATTCA GACAAGTGAA GGGACGACAA 3480

GCCAATATTT GAAACCAGAT AGCAGTTCTT ATAGTCAATT GAAATAAAAT CTGAAGAAAT 3540

CGAGTAGGAA ACTCATATCA ATGTTTAACA GTGTTCTATT CCAGATTCAT ACTCAATGAw 3600

AATTAAAGTG CAAACTAGGA AGTTAGCCGC AGGTGATACT TTGGGTACGG CA 3652

(2) INFORMATION FOR SEQ ID NO: 252:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 743 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

GTACCGTGGT GCCAAAGTAC AGCAAGGTTG GCTTTTTGAC AAACAATACC AATCTTGGTT 60

TTACATCAAA GAAAATGGAA ACTATGCTGA TAAAGAATGG ATTTTCGAGA ATGGTCACTA 120

TTATTATCTA AAATCCGGTG GCTACATGGC AGCCAATGAA TGGATTTGGG ATAAGGAATC 180

TTGGTTTTAT CTCAAATTTG ATGGGAAAAT GGCTGAAAAA GAATGGGTCT ACGATTCTCA 240

TAGTCAAGCT TGGTACTACT TCAAATCCGG TGGTTACATG ACAGCCAATG AATGGATTTG 300

GGATAAGGAA TCTTGGTTTT ATCTCAAATC TGATGGGAAA ATAGCTGAAA AAGAATGGGT 360

CTACGATTCT CATAGTCAAG CTTGGTACTA CTTCAAATCC GGTGGTTACA TGACAGCCAA 420

TGAATGGATT TGGGATAAGG AATCTTGGTT TTACCTCAAA TCTGATGGGA AAATAGCTGA 480

AAAAGAATGG GTCTACGATT CTCATAGTCA AGCTTGGTAC TACTTCAAAT CTGGTGGCTA 540

CATGGCGAAA AATGAGACAG TAGATGGTTA TCAGCTTGGA AGCGATGGTA AATGGCTTGG 600

AGGAAAAACT ACAAATGAAA ATGCTGCTTA CTATCAAGTA GTGCCTGTTA CAGCCAATGT 660

TTATGATTCA GATGGTGAAA AGCTTTCCTA TATATCGCAA AGTAGTGTCG TATGGCTAGA 720

TAAGGATAGA AAAAGTGATG ACA 743

(2) INFORMATION FOR SEQ ID NO: 253:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4010 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 253:

TTTTGGTTGA TGATACGAGG GATTTGGTGA TTCTTCTTGA CGATAGAAGT TTCAGCGACC 60

ATCATTTTTG AACAGTGATA GCACTTGAAT CGACGCTTTC TAAGGAGAAT TCTAGTAGGC 120

ATACCAGTCG TTTCAAGATA AGGAATTTTA GAAGGTTTTT GAAAGTCATA TTTCTTCAAT 180

TGGTTTCCGC ACTCAGGGCA AGATGGGGCG TCGTAGTCCA GTTTGGCGAT GATTTCCTTG 240

TGTGTATCCT TATTGATGAT GTCTAAAATC TGGATATTAG GGTCTTTAAT GTCTAGTAAT 300

TTTGTGATAA AATGTAATTG TTCCATATGA TTCTTTCTAA TGAGTTGTTT TGTCGCTTTT 360

CATTATAGGT CATATGGGAC TTTTTTTCTA CAATAAAATA GGCTCCATAA TATCTATAGT 420

GGATTTACCC ACTACAAATA TTATAGAACC GAATTAATTT AATTAGAGAG CCAACTTTCT 480

AATATAGTAA TCGCGTCATA ACAAGGTATC TATCATTCAT GGAGTTCCTC CTGTATACTA 540

TTAGTAAAGT AAAACTATTG GAGGATATTT TAATGCCACA ACCTATTGTT CCTGTAGAGA 600

TTCCACAATC TCGTCGTTTT GATTCTAAAA AGAGAAATGA TATTCTGCTT AAAATTCGTA 660

TTGGCAAGCT TGAAGTAAGT TTTTTTCAAT CTCTCAATCT CGAAATGGTA GAACAGCTTT 720

TGGATAAAGT GTTGCTCTAT GACAATTCAT CTATCTAGCC TAGGGCAGGT CTATCTCGTA 780

TGTGGGAAAA CGGATATGAG GCAAGGCATT GATTCATTGG CTTATCTGGT TAAAACCCAC 840

TTTGAATTAG ATCCTTTCTC CGGTCAAGTT TTTCTCTTTT GTGGTGGACG TAAAGACCGC 900

TTTAAAGCCC TTTACTGGGA TGGTCAAGGA TTTTGGCTAC TATATAAACG CTTTGAGAAC 960

GGAAAACTGA CTTGGCCCAG TACAGAAAAG GATGTCAAAG CTCTCACACC TGAACAAGTA 1020

GATTGGCTTA TGAAGGGCTT TTCTATCACT CCAAAAATAA ATTTATCAGA AAGTCGTGAT 1080

TTCTATTGAA ATGAGGACTT TCTTTTTAGT TATAATAAAG TTAGGAAATA AGGAGAGGAA 1140

GCCCATGGAA GAAGATTGAA AATCATTCAA CAACAGAGTG CTACAATTGA TAGTCTCACC 1200

AATGAACTTG CCCTTCTTCG TGAACAAGTG GCTTATCTAA CGCAAAAGCT CTATGGAAAA 1260

TCCTCTGAGA AAAGTGTTTG CCCATCTGGA CAACTCAGTC TTTTTGAAGA GGAACAAAAT 1320

ATGGAAGAAG ACTCTGACTT ACCCAGTTGA AAGAGAAGAA ATCACCTATA AACGTAAGAA 1380

AGCTAAAGGG AAACGTCAAG CTCTTCTTGC CCAATTTGAT TCAGAAGAAG TTCATCATCA 1440

AGTAGAAGAG AGCATTTGCC CTGATTGTCA GGGAGATCTA AAAGAGATTG GAGCAACCCT 1500

TCAACGACAA GAATTAGTCT TTATTCCTGC GCAATTAAAA CGAATAGATC ATATCCAACA 1560

CGCTTATAAG TGCCAAGCAT GCAGTGATAA AAATCCGAGT GATAAAATCG TGAAAGCTCC 1620

TATTCCTAAA GCCCCTTTGG CGCATAGCCT TGGCTCAGCT TCTATTATCG CTCACACCAT 1680

CCATCAGAAG TTTAATCTGA AGGTACCCAA TTATCGCCAA GAAGAAGATT GGGCTAAGAT 1740

GGGTTTACCA ATCACACGTA AGGAAATTGC TAATTGGCAT ATCAAGGCGA GTCAATACTA 1800

TTTGGAGCCC CTTTATAATC TTTTACGAGA AAAGTTGTTA GAACAAGCTC TTCTTCATGC 1860

GGATGAAACC TCTTATCGGG TTCTAGAGAG TGATAGTCAG TTGCCTTACT ATTGGACTTT 1920

TTTGTCTGGG AAAGCTGAGA ATCAAGCAAT CACGCTGTAC CACCATGATC AGCGTCGGAG 1980

TGGTTTAGTA GTACAAGAAT TCCTAGGAGA TTATTCTGGC TATGTTCATT GTGACATGTT 2040

GCGGCAGTAA CTTAGGACTT TAGTCCTCTA GTTCTGCCTA TGCGATAGCA GTCCAAGGTT 2100

TAGGAGTAAG GCGACGCTAA GCTTGGTAAA CTGCGAACAG CTAGAAGCTT ATCGTCAACT 2160

GGAAGAAGCT GCACTTGTTG GATGTTGGGC GCATGTGAGA AGGAAGTTTT TTGAAGTGCC 2220

CCCCAAGCAA GCAGATAAAT CATCCTTAGG AGCTAAAGGT TTAGCTTATT GTGATCAGTT 2280

ATTTTCCTTG GAAAGAGACT GGGAGGCTTT GCCAGCTGAT GAACGACTAC AGAAACGTCA 2340

AGAACATCTC CAGCCCCTAA TGGAAGACTT CTTTGCTTGG TGCCGCCGTC AGTCAGTTTT 2400

AGCAGGTTCA AAACTAGGAA GGGCAATTGA ATACAGCCTC AAGTATGAAG AAACCTTTAA 2460

GACTATTTTG AAAGACGGAC ATCTGGTCCT TTCCAATAAT CTAGCTGAAC GCGCCATTAA 2520

ATCATTGGTT ATGGGACGGA GTAAAAGAGT CCAGTGGACT CTTTTAGCCT GAGCTCAGTT 2580

TAAAAAAGCG AGGGTGGTTA TTTTCTCAAA GTTTTGAAGG AGCTAAAGCA AGAGCTATTG 2640

TTATGAGCTT GTTGGAAACA GCTAAACGTC ATCAATTATA GTGCGTTGAA TCTATAACAG 2700

TACGCATCGA CTGCTAAAAC ATTTCTATAA ATCAATTTTC CTTTCCTAAT CGATTTGTTC 2760

ATATCTTATT TCAATCCATT ATAAATAGCG AGAAATATCT ATCCTATCTT CTAGAATGTC 2820

TTCCAAACGA GGAAACTCTC GTAAACAAAG AGGTTTTAGA GGCCTATTTA CCGTGGACTA 2880

AAGTTGTACA AGAAAAGTGC AAATAAGAAA TCTCCAGATT AGGAACTATC CGTGAGTTCT 2940

CTAGTCTGGA GATTTTTCAA TAGACTTCGT TATTGGACGG TTACAATTTA TTATATGAAA 3000

ATCCCATATT ATTCTCCAAT TCTATATTTT ACCTTTCTAA ATGTATAGAT TAACTACCTA 3060

ATTATAGCAT ATAACGCAGA TTCCTTTCAA TCGTATGATT TACTGCATTA AATTAAGTAA 3120

AAAAATAAAG GCAGTCCGAA GACTGCCGAT ATTTATCTCT CATCTCTTTA ATTATGGTAA 3180

GTAAATAAAT AATTTCCCTA AAGATATGGA AATTATTAAT ACTATAAATA CATATTATAA 3240

AGTTTATAAA TACTGTAAAA ATCCTGAAGT TAATTTTCTA ATAAATATCA ATATGTGTTA 3300

GTATCTTTTA AATTTTTAGA CAATTTACTA GTTCTATAGA CATGTTTAAC AGACTCTATT 3360

TTACAATTCA AAAATTTCAT CTGCCACTTC ATTTAAAAAT TCTATATCAT GGGAAACAAT 3420

AAAAATTATT TTATCCATGG TTTTATACTT ATTAATCAGT TCAGATATTT TTATCATATT 3480

GGAATAATCC ATACCACTTG AAGGTTCGTC AAAAAAGACA AATGGAGAAT TCTTGCACAT 3540

AACAGATGCT ATTGCAAGCC TTTGCTTTTG CCCTCCTGAT AAACTCATCG GATGCCTTTC 3600

AATAAATTCG TCCAGGCATA AATCTTTTAA CCCAAATCAT TCATACCTCT CTCAACTAGA 3660

TGTAACTTAC AAAACCCCTG ACCTCATGAG CCACTTTCTT CCTCCTCATG AGGTCAGTTT 3720

TACTTTCTGC TGTTCCAGTA TCGTTTTTCC TCGCTAGATT TCCTCAAAAG GGCAGACTCC 3780

TCCCTTGGTT CGTCACACGA TTTTTTCATC TCGACTGTTC TTTAATGCAT CATTAACGAC 3840

GCTTTTCTTC TAGGTGGTTC ATAAGGAACA GGAAGATTCA GGTTGACTTT TCTAATCCTA 3900

GAATAAAGTG CTGAAAACAA TTCGGAATAG GCATAGAGAC TAGACAATTT GAGGAGCTGC 3960

TTGCGTCCTG TTCGAACACA TTTTCCCACC ACGTGAAGAA AAAGATGGCG 4010

(2) INFORMATION FOR SEQ ID NO: 254:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2789 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

ATGCATCCGT TTGTCAAGCC TAAATTGTAA TTTTTTTCAA TTTAAAACAG AAAAACCCAG 60

GAAAATGACA TAAAAATATC ATTCCTAGGC CTATTTATGC TATTTCTCTC TGAAAAATAT 120

GAGTATTCAG TCGGTCAAAT GAAGCTGAAC GAACTCATTT TCCCTCGCCT AATTCAATGA 180

TTCGATGACA TTGTTGGGCT ACATAAGCAT CGTGGGTCAC GATAATGACT GTTTTCCCCT 240

CTCGATTCAT CTCTAAGAGA AACTTCAAGA CCAAATCTCT ATTTTCAGGA TCCAGAGAAC 300

CTGTCGGTTC ATCGGCTAAA ATCAGCTGGC TGGGTTTTAA GATGGCTCTA GCAACTGCAA 360

TTCGTTGTTG TTCGCCCCCA GACAACTCGG AGACCCTTTG ATGCAAAGTA GCTGATAAAC 420

CTACTCTCTC TAAAATCTCT TCCACCTTTT TGAGCTTGTC TTTCTTAGGC AATTTCACAT 480

ATTTCAGCGC CACATGAGAT TGTACTCGAC CGTTTCATCA TCAATCAGGG CAAAATTTTG 540

AAACAGATAA GAGATATGTT CACGGATTAT TGTTTGCGAC TTAGCAGAAT TAACCGCTAG 600

ATTTGTCTGA CCAAAAATCT CATACCGTCC GCTATAATCA CCATCTATCA AACCCAATAA 660

ATTTAACAAG GTCGACTTCC CACTACCACT CTTACCAACA ATAGCTACCA AATCCCCCTG 720

ATCAATCCTG AGAGATAAGT TATCCAAAAT CACTTTTCCC CCAATGGTTT TGGTAATATT 780

TTTCAACTCA ATCATAAGAT GCCCCCTTTC AATAACTCTA CTAGACTTCT TTTCTCCATC 840

CTAGAAGCTA AGCCTAGCAC AAATAGTATA TCCAGACATG TAAAACCTGC AAACAGTAGA 900

AGTGGTAAGA ACGCATGGGC AAAGAAAATC AAGACTAGAA GAGGGAAACT ATAGCCCAGC 960

AAGAGCAGAA CGAGGAGAGG ACGGTAGCGA TCGACCAGTT TCCACCCCAT AAACTTCTTG 1020

GTAATGATAT CCCTGCGCTT CAATAAGAAA GTTGTTACTA GTAAGAAGTA GGAAATCATC 1080

ATGCTAAGGA GACCAAACAA AGCAAAGAGT AGGTTAAAAT TCCGAACAGC ATCTCGATAA 1140

GAATCCACTT TCTCTTGTTG AATGGCTTGA ATAGATGAAA ATTTTAAATA ATTTCCATCT 1200

GACAATTTCT CAACTAACTC TGTAATCTCT TTTTGATGTT GAACCGTATT TTCAATTTTA 1260

ATCGGATTAT TTAAGCCAGT TGTTGACAGG GAGGCTTTCT CATCCCACAT CATATCAGAA 1320

TCATTGACCA AGCTAATAAT TGGATTGGAG AGATTTTCCT TTCGCTTATC ACTATATGGG 1380

AAAAATGACC AATCTCCTTC ATAATAGGCA ATCTCGACAT CCATCTCCTC TATCGTTCGT 1440

1280

TTTTGCTGCT CTTCATACTT CATCGAATGA AAGGCAATTA ACTTCCCCAA GAGCTGATTT 1500

TTATCTTCTT CACCTTTCGT ACTTGCTGGC ATCAAAATAA CTTTTTTAAT ACCGGTATTT 1560

GGTAGCTTGA ATCCCTTGCT CTTTAGAAAA TTGCGATTGG CATAGTAAAC ATCCACCGTA 1620

TCTGTTAACT GATATTGCTG AATCTGTTCT GATTGGACAA AATTTTTTAC AGGAAGACTG 1680

CTACTCTGCA CATAGCCCGC CTGCGTTTTT TCTACCAAAT CCTGATAAAA TCGATAGAAA 1740

TAATCTGTAG ATTTCCCTGA CCCTGCTAGC TCTTCTTGCC ACAGATTATC ATTGAGTTTG 1800

AAGGTTTCTA AGGTCAGGTA ATTACCTTGA CTTACCCACT GTTGCTGATA AGCAAGTTCT 1860

TTGTTTTCTT GTTCTAAACT TCTGCCCACC CCAATCAGTA AGGCCGTCAG TAAAATAGTT 1920

GTCCCTATTT TCATCACATA ATTGAAGATA AGACCAAATT TGAAAGATGA AAAACCTTTC 1980

AGCAGAGAGC TGATTGTCAT TTTTTGGATT AAGAGGTAAG TCAACCAACT GATAAAGAGA 2040

TAAAGCTGCA ACAGCAAAAA ATGAGACAAC CACAGCATAG GAAACAAATC TTTTGGCTTA 2100

TAATCAAGCA AGAAAAACAC GCCTAGATTG ATCACAAGAG CCCCACCTAG GAGGAGGTAA 2160

AGGTTGCCTT TTACAACATC AGCTAAAACA GCCCTATCTT GAAAACCAAG TAATTTTTGT 2220

ACCCCAACTC TTTTCATCTC CATCATCGGT TGATACACTG TCACTAACAC AAGAAGCAAA 2280

ATAGCCAAGA CAAAAACAAT GGCAGATAAA AGCAAATCTC GATTTATGAC TTCCACTGCA 2340

CTTTTGTAGG TCGGCTCTAG CAAGGTAGCC TGGTCTATCT TGAAAAAATC GCTCCATTTC 2400

TGTACAATCC TATCCTTGTC CATCTCTTGT GTAGAAGTTA TCGTATAGCG ACCATTTAAA 2460

CTACGAGATG TATCCTTGAT ATAGGTTTGA AAAGTCATAA GCTGAATAGG TTTGGCTTTT 2520

AGAAAGGTCG GAATCGTACC AAGTTTATTG GAAATTTCTT TATTACTATA GACTCCTTCA 2580

CCATCTGTGG TAAAATCAAG AGAAGAAATC CCAAACTCTT GGTAGGGGAA GGTATCTTTA 2640

TCAAAAACAC CAGACTTGAC CACCTCATCA CCACTGTCTG TTTTGATGAT GGAGACTTTA 2700

TACTCCTTTG ATACATCCTC AAAAAATCGA AGAACAGACG CTGCAGGTTC GTTAATATCT 2760

TTCAAATACA AATCCAAAGA ATCTACAGG 2789

(2) INFORMATION FOR SEQ ID NO: 255:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2495 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 255:

CTGCGAATTT TATTAAAGAT AATGTGTTAA TTACAGCGGC TCACAACTAC TACAGACATG 60

1281

ACTATGGGAA AGAAGCGGAT GATATTTATG TTCTTCCGGC TGTTAGTCCA AGTCAAGAAC 120

CATTTGGAAA GATCAAAGTA AAGGAAGTTC GTTATTTGAA GGAATTTAGA AATTTAAATT 180

CTAAGGATGC AAGGGAATAT GACTTGGCTT TATTAATTCT AGAAGAGCCC ATTGGTGCAA 240

AATTAGGGAC TTTGGGTCTT CCTACTAGTC AAAAAAATTT GACAGGAATA ACTGTGACTA 300

TCACAGGCTA TCCATCATAT AATTTTAAAA TTCATCAAAT GTATACAGAT AAGAAACAAG 360

TTTTAAGTGA TGATGGCATG TTCTTGGATT ACCAAGTTGA TACTTTAGAG GGGTCTAGTG 420

GATCTACAGT TTATGATGCT AGTCACCGTG TAGTAGGAGT GCATACTTTA GGAGATGGAG 480

CTAATCAAAT TAACAGTGCA GTTAAATTAA ATGAACGAAA TTTGCCATTT ATTTAwTCGG 540

TTCTTAAAGG TTACTCTCTT GAAGGATGGA AGAAAATAAA TGGTAGTTGG TACCATTATA 600

GACAACATGA TAAACAAACG GGTTGGCAGG AGATAAATGA TACCTGGTAT TATTTAGACA 660

GTTCCGGTAA GATGCTTACA GATTGGCAAA AAGTCCATGG AAAATGGTAT TATCTCAATT 720

CAAATGGAGC AATGGTTACA GGTAGCCAAA CTATCGATGG TAAAGTTTAT AACTTCGCTT 780

CATCTGGTGA GTGGATTTAA TGTTGGAGGA TATATAAAAT GAAGCTTTTG AAAAAAATGA 840

TGCAAATCGC ACTAGCCACA TTTTTCTTCG GTTTGTTAGC GACAAATACA GTATTTGCAG 900

ATGATTCTGA AGGATGGCAG TTTGTCCAAG AAAATGGTAG AACCTACTAC AAAAAGGGGG 960

ATCTAAAAGA AACCTACTGG AGAGTGATAG ATGGGAAGTA CTATTATTTT GATCCTTTAT 1020

CCGGAGAGAT GGTTGTCGGC TGGCAATATA TACCTGCTCC ACACAAGGGG GTTACGATTG 1080

GTCCTTCTCC AAGAATAGAG ATTGCTCTTA GACCAGATTG GTTTTATTTT GGTCAAGATG 1140

GTGTATTACA AGAATTTGTT GGCAAGCAAG TTTTAGAAGC AAAAACTGCT ACGAATACCA 1200

ACAAACATCA TGGGGAAGAA TATGATAGCC AAGCAGAGAA ACGAGTCTAT TATTTTGAAG 1260

ATCAGCGTAG TTATCATACT TTAAAAACTG GTTGGATTTA TGAAGAGGGT CATTGGTATT 1320

ATTTACAGAA GGATGGTGGC TTTGATTCGC GCATCAACAG ATTGACGGTT GGAGAGCTAG 1380

CACGTGGTTG GGTTAAGGAT TACCCTCTTA CGTATGATGA AGAGAAGCTA AAAGCAGCTC 1440

CATGGTACTA TCTAAATCCA GCAACTGGCA TTATGCAAAC AGGTTGGCAA TATCTAGGTA 1500

ATAGATGGTA CTACCTCCAT TCGTCAGGAG CTATGGCAAC TGGCTGGTAT AAGGAAGGCT 1560

CAACTTGGTA CTATCTAGAT GCTGAAAATG GTGATATGAG AACTGGCTGG CAAAACCTTG 1620

GGAACAAATG GTACTATCTC CGTTCATCAG GAGCTATGGC AACTGGTTGG TATCAGGAAA 1680

GTTCGACTTG GTACTATCTA AATGCAAGTA ATGGAGATAT GAAAACAGGC TGGTTCCAAG 1740

TCAATGGTAA CTGGTACTAT GCCTATGATT CAGGTGCTTT AGCTGTTAAT ACCACAGTAG 1800

GTGGTTACTA CTTAAACTAT AATGGTGAAT GGGTTAAGTA ATGAAGGCTA ATTGTAAACT 1860

GTGATGGATA CTTAACTTTG TATAATAGGT GGATAAAAGT CTTCACAATC AAAAAACGCA 1920

TAGTATCAAG GTTTTTCTGT ACTGCCCTCA AACAGTTAGA CAATTAATTT ATCCGAAGgA 1980

TTTAGTTCTG TATTGCACAG GGCTAAGTCC TTTTAGTTTT ACCTTAATTC GTTTATTGTT 2040

GTAGTAATCA ATATAGTCTA TAATGGCTTG TTCCAATTGC TTAAGCGACT GAAACGACTT 2100

CTCATAACCG TAAAACATTT CCGATTTCAG AATCCCAAAG AAGGACTCCA TCATACTATT 2160

GTCTGGGCTG TTTCCCTTAC GTGACATGGA TGCTTGAATT CCCTTACTCT CTAGGAACCG 2220

ATGATAAGAA TCGTGTTGGT ATTGCCAGCC TTGGTCACTA TGGAGAATCG TATTCTCGTA 2280

GTGCTTCTCT GTGAATGCCT GTTCCAACAT TGTTTGTACT TGTTCTAAGT TGGGTGAAGT 2340

TGAAAGATTA TAGGCGATAA TTTCGCTATT AAAGCCATCT AAAACTGGTG ATAAGTAAAG 2400

CTTTTGAGTA CTTGCTGGAA TGGCAAATTC TGTCACATCT GTGTAGCACT TTTCCATTGT 2460

TTTAGAGCCT TCAAATTGGC CTTGAATGAG ATTCG 2495

(2) INFORMATION FOR SEQ ID NO: 256:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 870 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 256:

TACCACCGTA TTCATCCAGC AAGATTGCCA TTTGTCTTTG GGTATTTCGC AGTTCTTTTA 60

GCAAGTCATC CACAAAAATA GTTTCAGGTA CAAAAAGTGG ATCTTGTAAA ATTCTCTTCC 120

AAACAATATT GTCAAAACCG TCCACAAAGC CTGCCTTAAG GAGACTCTTG GTGTGAATGA 180

TTCCAATTAC ATTGTCCTTA TCCCCATCAT AAACCGGGAT ACGAGAATAA TTTTGTTTTA 240

AAATACTTTG GATAATGGCT TGACTATCAT CCTGAATATC CACCATAAAG GCATCCGTTC 300

GAGGAACCAT AACCTCTCGT GCCATCAGTT CATCGAGCGA AAAGACACCT TGTAGCATCT 360

CAATCTCATC AGCATCCAAT GTTTCTTCAC TATTTGTCAG CATATAGGCA ATTTCATCAC 420

GGGTCATCTT TTCATCCGCA TCATCGAATG ACATAGGAGT CAAATGGCTC AAGAAATTGG 480

TCGAAGCAGC TAAAAGCCAA ACAAAAGGAC TGACTAGTTT TCCGATCCCA ATGATAATCG 540

GCGCTGTACG AATTGCCAAG GCATCCTTTA GATTAAGAGC GATTCTCTTA GGATATAATT 600

CCCCAAAAAC GATGGAAATA TAGGTCAAAA ATGCCAAGGA TAGAAAAGTT GCCACGGCTT 660

GTGCTGTTTC GCCATTCCCA AGCCAAGAGG CAATCACACG TCCTAGAGTA TCAGTTAAAC 720

TCGCCCCTGA TAAGATTGTA ATCAGGGTGA TTCCTACCTG GATGGTTGAT AAAAAGTGGT 780

TAGGATTTTC TAGTACCTTC AGCAGGCGGA TGTAGCGTCT GTCTCCTTCT TCCGCCTTTT 840

GTTCAACTCG GGCACGATTA AGAGAAACGG 870

(2) INFORMATION FOR SEQ ID NO: 257:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1245 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 257:

CGTTCCCAGA AGCCCGCATT CTCATCGCCA ATGTCGTGAT TGATTTGGCC CTTTCTCCAA 60

AATCCAACTC AGCCTATGTA GCTATGGATA AGGCACTTGC TGACCTCAAA ACATCAGGGC 120

ACTTGCCTAT TCCGCGACAC CTGCGTGATG GGCACTACAG TGGAAGCAAG GAACTGGGGA 180

ATGCCCAAGA CTATCTCTAT CCACACAACT ATCCTGGAAA TTGGGTCAAG CAAGACTATC 240

TGCCAGAAAA AATTCGTAAT CATCACTATT TCCAAGCAGA AGATACTGGT AAATATGAAC 300

GGGCTTTGGC TCAAAGAAAG GAAGCTATCG ACCGTTTGCG AAAAATCTGA AATCCTTTTC 360

AAAAAATTGC ACTTTCCTCT TGATTTTTTT TGAAAAAGTG GTATCATATA AATATAGAAA 420

CGCTGTGGTG TACGACTTCA CACTTAAGTG TTGACCGACT ATTTTTTGTA TTATTAGGGA 480

AACAAAAGTC TTCTAACAGC ATGTAGGCCG TCTCACACGG AAACAGCTTC AGTTAGAGCG 540

AGTTGCCCAC CTGCTTAATT GCGCGGGTTC AATACAAACC GTGAAGTTTC GGCACCAATA 600

CAGCTTTTTT CTTTGCCTCC TTAGCTCAGC TGGCAGAGCA GCGGACTCTT AATCCGTGGG 660

TCACAGGTTC GATCCCTGTA GGGGGCATAT AAATACAACA GGAAAAGCCT TATAATATAG 720

GGCTTTTTTT GCTTTCCTTT TAAAAATTGT CGTGCAATTT GCCGTGTTTT TACAACAAAC 780

TTTTCACAGC CATAAACTCC TCACTAATTT TTTCCTCCAA GGTATGCCCA TAAACGTCAA 840

TCAACATGGA GATATCTTTA TGTCCTAAAA TTTGGCTCTT TGTCAACTGT AGTGGGTTGA 900

AGTCAGCTAA GCTCGAGAAA GGACAAATTT TGTCCTTTCT TTTTTGATAT TCAGAGCGAT 960

AAAAATCCGT TTTTTGAAGT TTTCAAAGTT CCGAAAACCA AAGGCATTGC GCTTGATAAG 1020

TTTGATGAGA TTATTGGTCG CTTCCAATTT GGCGTTAGAA TAGTGTAGTT GAAGGGCGTT 1080

GACGATTTTC TCTTTGTCCT TTAGAAAGGT TTTAAAGACA GTCTGAAAAA GAGGAGGAAC 1140

CTGCTTTAGA TTGTCCTCAA TGAGTCCGAA AAATTTCTCC GGTGCCTTAT TCTGAAAGTG 1200

1284
AAACAGCAAG AGTTGATAGA GCTGATAGTG ATGTTTCAAG TCTTG 1245

(2) INFORMATION FOR SEQ ID NO: 258:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1684 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 258:

ATGCCTATGT AACTCCACAT ATGACCCATA GCCACTGGAT TAAAAAAGAT AGTTTGTCTG 60

AAGCTGAGAG AGCGGCAcCC AGGCTTATGC TAAAGAGAAA GGTTTGACCC CTCCTTCGAC 120

AGACCATCAG GATTCAGGAA ATACTGAGGC AAAAGGAGCA GAAGCTATCT ACAACCGCGT 180

GAAAGCAGCT AAGAAGGTGC CACTTGATCG TATGCCTTAC AATCTTCAAT ATACTGTAGA 240

AGTCAAAAAC GGTAGTTTAA TCATACCTCA TTATGACCAT TACCATAACA TCAAATTTGA 300

GTGGTTTGAC GAAGGCCTTT ATGAGGCACC TAAGGGGTAT ACTCTTGAGG ATCTTTTGGC 360

GACTGTCAAG TACTATGTCG AACATCCAAA CGAACGTCCG CATTCAGATA ATGGTTTTGG 420

TAACGCTAGC GACCATGTTC AAAGAAACAA AAATGGTCAA GCTGATACCA ATCAAACGGA 480

AAAACCAAGC GAGGAGAAAC CTCAGACAGA AAAACCTGAG GAAGAAACCC CTCGAGAAGA 540

GAAACCGCAA AGCGAGAAAC CAGAGTCTCC AAAACCAACA GAGGAACCAG AAGAATCACC 600

AGAGGAATCA GAAGAACCTC AGGTCGAGAC TGAAAAGGTT GAAGAAAAAC TGAGAGAGGC 660

TGAAGATTTA CTTGGAAAAA TCCAGGATCC AATTATCAAG TCCAATGCCA AAGAGACTCT 720

CACAGGATTA AAAAATAATT TACTATTTGG CACCCAGGAC AACAATACTA TTATGGCAGA 780

AGCTGAAAAA CTATTGGCTT TATTAAAGGA GAGTAAGTAA AGGTAGCAGC ATTTTCTAAC 840

TCCTAAAAAC AGGATAGGAG AACGGGAAAA CGAAAAATGA GAGCAGAATG TGAGTTCTAG 900

TTCTCATTTT TTTCATGAAA ATGTGCAAAA TATAGTAGAT TGAAACTAGA ATAGTATACC 960

TCTACTTCTA AAACATTGTT AGAAATCGAT TTGACTGTCC TGTTCTTATT TCATTTTACT 1020

ATATCTTAAC AGATAGTGTA AATAAAGATA AACTATTTAC TGGCTAATTA ATCAGTTAAA 1080

CACTAGTTAA GGAGTAATGA TGAAAAAAAG AACAATACTA TTATTGATGG CCAGTCTGTT 1140

AGCTCTTGTC TTAGGAGCAT GTGGTTTCTT GGACATATTG ATCCTGGATC ATTCTCATCA 1200

GGATTACTCT TTACTGCTAT TTTAGAAACT GGGGTGGTTT GATGGAAAGT ATTGGTCTTG 1260

TTATCGTTTC ACATTCCAAA CACATTGCAG AAGGTGTTGT TGAACTGATT AGTAAAGTAG 1320

CTAAAGATGT TCCGATTACT TATGTAAGAG GAACCGAGGG CGGAGGAATT GGAACGAGTT 1380

TTGAACAAGT AGATAGGGTT GTTTCCGAAA ATCCAGCAGA TACTTTACTT GCCTTTTTTG 1440

ACCTAGGTTC TGCTAAAATG AACTTAAAAA TGGTGACTGA TTTCAGTGAT AAAAGTATCA 1500

TCATCAACAG GGTTCCAATT GTAGAAGGTG CCTATAATGC AGCTGCTCTT CTTCAGGCTG 1560

GTGCAGAACT GTCAGTTATT CAAACACAGT TaGCGGAgCt TGAAATCAAT AAATAAGGAA 1620

TTTTACTATA ACTCTTTTTA TAGATAAGCT ATTGaTTATC TCAACTATAA TAATGTTAAG 1680

TnAA 1684

(2) INFORMATION FOR SEQ ID NO: 259:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 970 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 259:

AGGAGTGGAG AnATATGAAG ACACAAATTT TCACATTATT GAAAATCGTT GCTGAGATTA 60

TTATTATTTT GCCATTTCTA ACTAATCTAT AAGTTCTTTA TATTGCTGAA AACGCAATTC 120

AAAAAGGGCT ATTAATTGTG GATTTTCTAA TACCTGCAGA GATTGGATAA AGCGTTCAAT 180

CTCTTTTTGA TTGCTTCCCT TTGTTTGAAG AAAGACACTC ATCTTCTTTA AAAATTGCCA 240

CGATACTTTT TCAAAAACAT CATACGGTCG TAACATCCTC TCCAACTCGG CTTCGAAGAT 300

TGGGATGTAG GAGAAAAGTT TTCGCTCCAT GAGTTCTGAT AAGATATTTA AGAGTCCTTG 360

CTTCATATAC AATCGATTGT GTACTAACTC TTTAAATTCT TTGGATTTTT CGAGTAAGGA 420

GGTTGATAAA AAAATCAGAT CTTGATTGCT CAAGAAGGGC ATGGTATTGC AAAAGAGATA 480

GAGTTCAAAC CAGGTCCAAG ACTCGATAGC ATAGAGATAG GTGGTCAAAA ACTCGCTATC 540

CTCCTCTGCT AGTGGGTAGC TTTTATTTAG TGAATGGATG GCATCTTTAA TCACGATGGC 600

ATTCAAACGA CGATAGGTCT GCGCCATCTG TTCTTGATCG ACTTCCTCCA ATAGCTGCTC 660

TAAAGCAGCT ATATCCTGAT GGGCAAAGCG ATTCACAACC TTTCGACCGA TTCGCATATG 720

TGGAGATTCT TGATAGTTGT TGAGCTTGTG CCCAAACTCA TCAAAGGTCA CATTTATACC 780

TTGGATAGCT AGAATCAACT TATCCGCAGA CAGCATAGAC TGCCCTAGTT CAAACTTGGA 840

CAACTGAGAA GCTGTTAGAC CCTCACAAGC CACATCTGAC TGCTTGAGCT TTCTCGCCAA 900

ACGTAATTCC TTGTAAAATT CCCCCAGTTC CATTCTCTCA ATCATCTGAC CACCTCCTAG 960

CTTTTGCAGG 970

(2) INFORMATION FOR SEQ ID NO: 260:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2996 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 260:

GTTGACCACG GGTAAAACTA CCCTAACTGC AGCTATCACA ACTGTTTTGG CACGTCGCTT 60

GCCTTCATCA GTTAACCAAC CTAAAGACTA TGCGTCTATC GATGCTGCTC CAGAAGAACG 120

CGAACGCGGT ATCACTATCA ACACTGCGCA CGTTGAGTAC GAAACTGAAA AACGTCACTA 180

CGCTCACATC GACGCTCCAG GACACGCGGA CTACGTTAAA AACATGATCA CTGGTGCTGC 240

TCAAATGGAC GGAGCTATCC TTGTAGTAGC TTCAACTGAC GGACCAATGC CACAAACTCG 300

TGAGCACATC CTTCTTTCAC GTCAGGTTGG TGTTAAACAC CTTATCGTCT TCATGAACAA 360

AGTTGACTTG GTTGACGACG AAGAATTGCT TGAATTGGTT GAAATGGAAA TCCGTGACCT 420

ATTGTCAGAA TACGACTTCC CAGGTGACGA TCTTCCAGTT ATCCAAGGTT CAGCACTTAA 480

AGCTCTTGAA GGTGACTCTA AATACGAAGA CATCGTTATG GAATTGATGA ACACAGTTGA 540

TGAGTATATC CCAGAACCAG AACGTGACAC TGACAAACCA TTGCTTCTTC CAGTCGAGGA 600

CGTATTCTCA ATCACTGGAC GTGGTACAGT TGCTTCAGGA CGTATCGACC GTGGTATCGT 660

TAAAGTCAAC GACGAAATCG AAATCGTTGG TATCAAAGAA GAAACTCAAA AAGCAGTTGT 720

TACTGGTGTT GAAATGTTCC GTAAACAACT TGACGAAGGT CTTGCTGGAG ATAACGTAGG 780

TGTCCTTCTT CGTGGTGTTC AACGTGATGA AATCGAACGT GGACAAGTTA TCGCTAAACC 840

AGGTTCAATC AACCCACACA CTAAATTCAA AGGTGAAGTC TACATCCTTA CTAAAGAAGA 900

AGGTGGACGT CACACTCCAT TCTTCAACAA CTACCGTCCA CAATTCTACT TCCGTACTAC 960

TGACGTTACA GGTTCAATCG AACTTCCAGC AGGTACTGAA ATGGTAATGC CTGGTGATAA 1020

CGTGACAATC GACGTTGAGT TGATTCACCC AATCGCCGTA GAACAAGGTA CTACATTCTC 1080

TATCCGTGAG GGTGGACGTA CTGTTGGTTC AGGTATGGTT ACAGAAATCG AAGCTTAATT 1140

CGATTTAGTT CCCAGAAGAA CAATTATTTA AGTTAGACAC TAAAAGAATC TTGCTTGGCA 1200

AGGTTCTTTT TTTAGATATT GAACTAATAC TCAATGAAAA TCAAAGAGCA AACTATAATA 1260

TATTGAAACT AGAATAGTAC ACATCTACTT CTAAAACATT GTTAGAAATC GATTTGACTG 1320

TCCTGATCGA TTTGTCTTGT TCTTATTTCA TTTTACTATA GAAAGTTAGC TACAGACTGC 1380

TCAAAACATT GTTTTTAGGT TGTAGATAGA ACTGACGAAG TCAGtAACAT CTATACGACA 1440

AGGCGAAGCT GACGCGGTTT GAAGAGATTT TCGAAGAGTA TAATACTAGA CTAAAATCAA 1500

AAAGCATTAT ACAATAGTAA TATGAAATCA ATTAAAGAAG AAATCCAAAC CATCAAAACA 1560

CTTTTAAAAG ACTCTCGTAC AGCTAAATAT CATAAACGCC TTCAAATCGT TCTATTTCGT 1620

CTGATGGGCA AATCTTATAA AGAGATTATA GAACTTTTAT AGTGGTTTGA AATAAGATGT 1680

GAACAACTCT ATCAGGAAAG TCAAACTAAT TTATAGAAAT ATTTTAGCAG CCAAGGTGTA 1740

CTGTTATAGA TTCAATACAC TTTAGACTGT AATCAAACAA CGATTTGGCG AAATGTAAAA 1800

AATATGAGGA GTTCGGACTC GACTCTCTCC TTCAAGAAAC ACGTGGTGGT CGTAACCATG 1860

CTTATATGAC GGTTGAGCAA GAGAAAGTCT TTCTTGCCCG CCATTTGAAG GCTACAGAGG 1920

CAGGAGAATT TGTTACAATT GATGCCTTAT TTCAGGCTTA TAAAAAGGAG TTAGGTCGTT 1980

CCTACACACG TGATGCCTTC TATCAACTGT TGAAGCGCCA TGGTTGGCGA AATATTACGC 2040

CACGTCCAGA ACATCCTAAG AAAGCAGATG CTCAAACCAT TGTCGCGTCT AAAAATAAAG 2100

TCTCAATTCA AGAAGACAAG TGAACTGCAC CCCAAAAGTT AGACAGAAAA AATCTAACTT 2160

TTGGGGTGTT TTTATTATGA AATTAACTTA TGATGATAAA GTTCAGATCT ATGAACTTAG 2220

AAAACAAGGA TATAGCTTAG AGAAGCTTTC AAATAAATTT GGGATAAACA ATTCTAATCT 2280

TAGGTACATG ATTAAATTGA TTGATCGTTA CGGAATAGAG TTCGTCAAAA AAGGAAAAAA 2340

TCGTTACTAT TCTCCTGATT TAAAACAAGA AATGATTCAT AAAGTCTGAC ATGAAGGCTG 2400

GACTAAAGAT AGAGTTTCTC TTGAATACTG TCTCCCAAGT CGTACGATAC TTCTTAACTG 2460

GCTAGCACAA TACAGGAAAA ACGGGTATAC TATTGTTGAG AAAACAAGAG GGAGAGTACC 2520

TGAGAGCGGA GAATGCCATC CTAAAAAAGT TAAGAGAACT CCGATTGAAG GAGGAAAAAG 2580

AGAAAGAAGA AAGACAGAAA TTATTCAAGA ATTAATGACT GAGTTTTCGT TAGATATTCT 2640

TCTAAAAGCC ATTAAACTAG CTCGTTTGAC CTACTACTAT CACTTGAAAC AGCTAGATAA 2700

ACCAGATAAG GACCAAGAGC TTAAAGCTGA AATTCAATCC ATTTTTATCG AACACAAGGG 2760

AAATTATGCT TATCGTCGGA TTTATTTAGA ACTAAGAAAT CGTGGTTATC TGGTAAATCA 2820

TAAAAGAGTT CAAGGCTTGA TAAAAGTACT CAATTTACAA GCTAAAATGC GACAGAAACG 2880

AAAATATTCT TCTCATAAAG GAGACGTTGG CAAGAAGGCA GAGAATCTCA TTCAAGGACA 2940

ATTTGAAGGC TCTAAAACAA TGGAAAAGTG CTACACAGAT GTGACAGAAT TTGCCG 2996

(2) INFORMATION FOR SEQ ID NO: 261:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 837 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 261:

CTTATCAACT CCCGACATGG CTCTCAGACC AATCCAAATC CCTAAAAAAA TCAGAACAAG 60

GATGGTGGTC AAGATCAAAC TCTCGAAATA TAAAGAAAAT AGTTGCAGTA GCATGATTTC 120

TCTCATTTCT ATCTTTTTTA AAGAGTAAAC TCAGCTAGTC CAACTAACTG AGTTTTCCTT 180

TATCTATTAT ATCAAATATA AGTCCGTTTG TAACTAGCGA AGAATTCTTT TGTCCGCTCT 240

TCTTTAGGGG TGTGGATAAT CTCATCCGGA GTTCCAGACT CGATGATTTT CCCCTTATCT 300

AAGAAGAGAA TTTTATCCGC AACTTGGGCT ACAAAGGACA TGTCATGACT GACCAAAATC 360

ATGGTCTGAC CTGACTTAGC AGCATCTGCA ATAGACTTTT CTACTTCACC GACCAATTCT 420

GGGTCAAGGG CTGAAGTTGG TTCGTCTAAG AGCAAAACAT CTGGTTTCAT AGCAAGCGCA 480

CGCGCTAGGG CAACCCGTTG CTTCTGTCCA CCTGATAAAT GGCGAGGATA ATGGTTTTCA 540

CGGTCCGAAA GCCCAACCTT AGCCAACTCT TCCTTGGCAA TCTTAGTCGC TTCTTGGTCA 600

GATAATTTCT TGACAACAAC CAAGCCTTCT TTCACATTAT CAAGTGCTGT TCGGCGTTCA 660

AACAAATTAA ACTGTTGGAA AACCATAGAC AACTTACGAC GTAGGGCAAG GATTTCTTCT 720

TGAGTGATTT TAGAAAAATC AACTGAAAAA CCATCAATCT GAATAGAGCC ACTGTCAGGT 780

GTTTCTAGAT AATTGAGACT GCGAGAAAGG TTGATTTTCA GCTCTGAAGA CCAATCA 837

(2) INFORMATION FOR SEQ ID NO: 262:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 868 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

CCGAACAAAA TGGGCTAATT AGATTATAGT AAGAAAGGTA AGTTAAAAAT GAGAATTGCA 60

ATTGGATGTG ACCACATCGT AACTGATGAA AAAATGGCGG TTTCAGAATT TTTGAAATCA 120

AAAGGATATG AAGTCATTGA CTTTGGTACC TATGACCATA CACGGACTCA CTACCCAATC 180

TTTGGTAAAA AAGTAGGGGA AGCTGTAACT AGCGGTCAAG CTGATCTTGG AGTATGTATC 240

TGTGGTACTG GTGTTGGTAT CAACAACGCT GTAAATAAAG TTCCAGGTGT TCGTTCTGCC 300

TTGGTTCGTG ATATGACAAC AGCCCTTTAT GCTAAAGAAC AATTGAACGC TAACGTTATT 360

GGTTTTGGTG GTAAAATTAC TGGTGAATTG CTTATGTGTG ATATCATCGA AGCTTTCATC 420

CATGCTGAAT ACAAACCAAC TGAAGAAAAC AAAAAATTGA TTGCGAAAAT TGAACATGTT 480

GAAAGTCACA ATGCTCAACA AACAGACGCA AACTTCTTTA CAGAATTCCT TGAGAAATGG 540

GATCGTGGAG AATACCACGA CTAAGAGGTG ACCTATGATT TTAACAGTCA CAATGAACCC 600

ATCCATCGAT ATTTCCTATC CCTTGGATGA GTTGAAGATT GATACTGTCA ATCGTGTGGT 660

GGATGTAACC AAAACGGCTG GTGGTAAGGG ACTCAATGTT ACCCGAGTAC TTTCAGAATT 720

TGGCGATTCT GTTCTTGCTA CTGGTTTAGT GGGTGGCAAA CTTGGTGAGT TTTTGGTTGA 780

ACATATCGAT AATCAAGTAA AGAAAGATTT CTTCTCAATT AAGGGAGAAA CTCGTAACTG 840

TATCGCTATT CTCCACGGAG ACAACCAA 868

(2) INFORMATION FOR SEQ ID NO: 263:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3744 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

CCGTTCAAAG TCTTCATAAG ACTCGAAAGT CACAGTTCTT TCGTTCTTGC TGGCATCTAT 60

ATAGGTAATT TCAATCATGT TTAAAACTCC TTTGTTTAAT GCTAACTTTA TTTTACTCCT 120

TATAAAAGAG AATGTCAAGA AAAATGATTG CGCACGCAAC TTTTTTTAAA ATCATCTTAA 180

ATCAAGAAAT CCAAACCTGC TTCCAAGCTT TCTTCGACAG TCTTTTGTAG CGAGGCCAGT 240

GTCTTTTGCC CATCATTTGT CAGGCAGATA AAACTAGAGC GTCTATCTTG ATGGCAACAC 300

ATGCGACTGA GTAGACCGCA ATTTTTAGCT TCCAAGCGAG CCACCATCCT AGAAACTGCG 360

CTCGGGCTCA GATGAAGCTT ATCTGGCAGG TCAATCTGGC GTAGAGATTT TTCTTCAGCC 420

AAGTCCAGAT AGTAGAGCAG GTAGAACTCT TTCAAGGTCA GACTTTGCTC GCTCTGTTGG 480

GCAATGGTCT CTTCCAAGAG ACTTTCAATT TCTTTCTGAC GCCGATTGAA GTCAAACCAT 540

TTTTCCAAAT AGGTCATAGT GTCTCCTTTC TTTTTAGAGT CATAAATAGA AGAAAGTCCA 600

TTAACGGGCA GTCTCTGCGT CACAAGATGA TTGCGCATGC AATAATTATA CTACTTTTCA 660

AGAATGCTGG CAAGCTCTGT TTTTTAGTGG TTTTATTTTT GTGTGAATAA TGGGGGAATC 720

CTATTGTTTC AATTTCTAAC TCCTTATCAC ATTCGAATTC AGATTTTATT TCATTTCTCT 780

ATCTATAGTT GCTTAGTTTA AAATAAGCAT GGTCTAATAA AGCTATGCAT ATAGTACTGA 840

TTTTAAACAA GGAGCATTAG ATTCCATTAA AGGAGGGCAC AGACATGTCG AGGCGGCCAA 900

1290

AGTTTTTGAT GTCGGCGTCA GAACTCTCTT CACGTGGGAA AAGAAAGACG TAAACAAGGG 960

AACTTAGAGC GGAAAAAGCG AGTCGTCAAA AAGCGTAAGA TCCCTTTAGA AGAATTGAAA 1020

GCCTTTGTAG AGGCTCATCC AGACGCTTTT TTACGGGAAA TTGCGGCCCG TTTTGATTGT 1080

GCTTTGCCCT CCGTATGGGC AGTTTTAAAG CAGATTAAGG TCATTTTAAA AAAGACGACC 1140

AGTTTTAGGG AACAAAAGCC TGAGAAAGTT TCTGAGTTTC TTGATATTTT GGATAACCTA 1200

AAAGATTTAC CAGTCCTATA TATTGACGAA ACGGGAATCG ACCGCTACCT CTATCGTCCT 1260

TATGCAGGGG CTCCTAGAGG GGAGAAAGTC TATGGCAAGA TTAGCGGACG GCGTTTTGAG 1320

CGGACTAATG AGGTGGAGCA AAAACTCAAT GGTAGTTTTC TAATCAGATA TATTGATTCA 1380

CAAATTAGAG AATGAAAGAA TAATTATGCA TAAAAATAGG AATATAAACC AAAAATTAGC 1440

TGATTTATAC TCATTTGCGT GTCTTTATAA AAAACTTATC TTATAATATA TATATATATA 1500

TATACAAAAT AGTAAAATGC TTTTTTTTTT TAGCAAAAAT ACCTCAAGTT TCTTGCTATT 1560

TTGGGTTCCC TATTCTATAA TTATAGTATG GTAATTTATT TATATCCATA CATGAAAATA 1620

ATACTCGAAA GGAAATTTCA AAATATTTTT TAGACGTCAG AAGGGTGAAT ATAGAGAAAC 1680

AGACCGAGTA ACTCGGTTCA AATTAATCAA ATCAGGGAAG CATTGGCTAC GGGCCTCGAC 1740

TTCTCTTTTT GGCTTGTTTA AGGTCTTGCG AGGTGGTGTT GATACTACTC AGGTCATGAC 1800

CGAAACGGTA GAAGATAAAG TAAGTCATTC AATTACTGGG CTTGATATCC TCAAGGGGAT 1860

AGTTGCTGCG GGAGCTGTCA TAAGTGGAAC CGTTGCAACT CAAACGAAGG TATTTACAAA 1920

TGAGTCAGCA GTACTTGAAA AAACTGTAGA GAAAACGGAT GCTTTGGCAA CAAATGATAC 1980

AGTAGTTCTA GGTACGATAT CTACAAGTAA TTCAGCGAGT TCAACTAGTT TGTCAGCTTC 2040

AGAGTCGGCA AGTACATCTG CATCTGAGTC AGCCTCAACC AGCGCTTCGA CCTCAGCAAG 2100

TACAAGTGCA TCAGAATCAG CAAGTACATC GGCTTCGACA AGTATTTCTG CATCATCTAC 2160

TGTGGTAGGT TCACAAACAG CTGCCGCTAC AGAAGCAACT GCTAAGAAGG TCGAAGAAGA 2220

TCGTAAGAAA CCAGCTAGTG ATTATGTAGC ATCAGTTACA AATGTCAATC TCCAATCTTA 2280

TGCTAAGCGA CGCAAGCGTT CAGTGGATTC CATCGAGCAA TTGCTGGCTT CTATAAAAAA 2340

TGCTGCTGTT TTTTCTGGCA ATACGATTGT AAATGGCGCC CCTGCAATTA ATGCAAGTCT 2400

AAACATTGCT AAAAGTGAGA CAAAAGTTTA TACAGGTGAA GGTGTAGATT CGGTATATCG 2460

TGTTCCAATT TACTATAAAT TGAAAGTGAC AAATGATGGT TCAAAATTGA CCTTTACCTA 2520

TACGGTTACG TATGTGAATC CTAAAACAAA TGATCTTGGT AATATATCAA GTATGCGTCC 2580

TGGATATTCT ATCTATAATT CAGGTACTTC AACACAAACA ATGTTAACCC TTGGCAGTGA 2640

TCTTGGTAAA CCTTCAGGTG TAAAGAACTA CATTACTGAC AAAAATGGTA GACAGGTTCT 2700

ATCCTATAAT ACATCTACAA TGACGACGCA GGGTAGTGGG TATACTTGGG GAAATGGTGC 2760

CCAAATGAAT GGTTTCTTTG CTAAGAAAGG ATATGGATTA ACATCATCTT GGACTGTACC 2820

AATTACTGGA ACGGATACAT CCTTTACATT TACCCCTTAC GCTGCTAGAA CAGATAGAAT 2880

TGGAATTAAC TACTTCAATG GTGGAGGAAA GGTAGTTGAA TCTAGCACGA CCAGTCAGTC 2940

ACTTTCACAG TCTAAGTCAC TCTCAGTAAG TGCTAGTCAA AGCGCCTCAG CTTCAGCATC 3000

AACAAGTGCG TCGGCTTCAG CATCAACCAG TGCCTCGGCT TCAGCGTCAA CCAGTGCGTC 3060

AGCTTCAGCA AGTACCAGTG CTTCAGTCTC AGCATCAACA AGTGCTTCAG CCTCAGCATC 3120

GACAAGTGCC TCGGCTTCAG CAAGCACATC AGCATCTGAA TCAGCGTCAA CCAGTGCTTC 3180

GGCTTCAGCA AGTACCAGTG CTTCAGCTTC AGCATCAACC AGCGCCTCGG CCTCAGCAAG 3240

CACCTCAGCT TCTGAATCGG CCTCAACCAG CGCCTCGGCC TCAGCAAGCA CCTCAGCTTC 3300

TGAATCGGCC TCAACCAGCG CCTCAGCCTC AGCATCAACG AGTGCTTCGG CTTCAGCAAG 3360

CACAAGCGCC TCGGGTTCAG CATCAACGAG TACGTCAGCT TCAGCGTCAA CCAGTGCTTC 3420

AGCCTCAGCA TCAACAAGTG CGTCAGCTCA GCAAGTATCT CAGCGTCTGA ATCGGCATCA 3480

ACGAGTGCGT CTGAGTCAGC ATCAACGAGT ACGTCAGCCT CAGCAAGCAC CTCAGCTTCT 3540

GAATCGGCCT CAACCAGTGC GTCACCTCAG CATCGACAAG CGCCTCAGCT TCAGCAAGTA 3600

CCAGTGCTTC AGCCTCAGCG TCGACAAGTG CGTCGGCCTC AACCAGTGCA TCTGAATCGG 3660

CATCAACCAG TGCGTCAGCC TCAGCAAGTA CTAGTGCATC GGCTTCAGCA TCAACCAGTG 3720

CCTCGGCTTC AGCGTCAAAC AGTG 3744

(2) INFORMATION FOR SEQ ID NO: 264:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 795 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

CGATAAAGAG GCCTTGAGTA ATCTCAATTT GCAGATTGAA AATGGAGAGA TTATGGGCTT 60

GATTGGTCAT AATGGGGCTG GAAAATCGAC CACTATAAAA TCCCTAGTCA GTATCATTTC 120

ACCCAGCAGT GGTCGTATTT TGGTAGACGG TCAGGAGTTA TCGGAAAATC GCTTGGCTAT 180

TAAACGAAAG ATTGGCTACG TAGCAGACTC GCCTGACTTA TTTTTACGCT TAACGGCCAA 240

TGAATTTTGG GAATTGATCG CCTCATCCTA TGATCTGAGT AGATCTGACT TGGAGGCTAG 300

TCTAGCTAGG CTATTGAACG TTTTTGATTT TGCTGAAAAT CGCTATCAGG TTATTGAAAC 360

TCTTTCTCAC GGAATGCGTC AGAAAGTCTT TGTCATCGGA GCACTCTTGT CTGATCCCGA 420

TATTTGGGTC TTGGATGAAC CCTTGACTGG TTTGGATCCC CAGGCTGCCT TTGATTTGAA 480

ACAGATGATG AAGGAACATG CACAAAAAGG GAAGACAGTC TTGTTTTCAA CTCATGTCCT 540

AGAGGTGGCA GAGCAAGTCT GTGATCGGAT TGCCATTTTG AAAAAGGGGC ATTTGATTTA 600

TTGTGGTAGT GTAGAGGACT TGAGAAAAGA TTACCCAGAC CAGTCTTTGG AAAGTATCTA 660

CCTTAGTCTT GCTGGTAGAA AAGAGGAGGT TGCGGATGCG TCTCAAGGTC ATTAAAAAAT 720

TAGTTGATAT CAATATCCTT TATTCATCTC AAGAAGCTAA TCTGGCTAAT CTACGAAAGA 780

AGCAGGCTAA GAATC 795

(2) INFORMATION FOR SEQ ID NO: 265:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2231 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

TGGTAATGTG CTTGGCAGCw TCCTTGACAC TGCTACTACC ATTTCCCATA GCGACCGACA 60

TACCAACGCC AGCCAGCATT TCAAGATCAT TATCTGAGTC ACCAAAAGCC ATGACTTGGT 120

TGAGGTCAAA GCCATATTCT TTCCCAACTC GGCGAATGCC TTCTAATTTA GAATTTCCCT 180

GATTGATGAC ATCCGATGCA AAAGGATTGC TACGTGTCAA TTTCAAGTCT TCAAAATCAG 240

CTGCCGCCTT CTCAGATTCT TCTGGTGTCA TCAGCATCAA AACTTGGTAG ATAGGCTGAT 300

TCATCAGGTG AAGCAGGTCC TCTTCCTTTT GGGGAACAAC CTTGCTGACC ATGCGATTAA 360

AAGACTGACT CACCGTCCGA GTTAAAACAG AGGGAACGAA GCGACTAATT CGTTGGGAAA 420

AAGAACCCAG ACCAAAGGAC ATGATTTTAG AACCCAACAT GGCATCCTTG GTCCCTAGAG 480

CAATCTCCGT GCCCTCTTTT TTAGCATAGC TAATTAGATG GCGCAAATGT AACTTGGAAA 540

TAGGGCTCGT GAACAAGACT CTGTCTTTAC TAAAGATATA CTGGCCATTA TAGGTTACCG 600

CAAAATCCAG ATCCAAATCG TCCATCAATT CCTTAACAAA AAAAGGTCCT CGCCCTGTCG 660

CTACGCCAAC TAGTACCCCT TGTTCTTTGA CAATCTTAAT CGCATCCTTA GTGGATTTCA 720

AAACACTCTT GCGATTGTTG ACCAAGGTTC CATCGATATC AAAAAAAACA GCTTTGACTT 780

CCATCCTATC CCAATCTCCC CTTTTGTGAT ACAATGATTA TACCACATTT CAGAAAGAGT 840

GAGTAAATCA TGCCTAAGAA AATCCTTGTT TTACATACGG GTGGAACTAT TTCCATGCAG 900

AACCGCCGAC ATTCAAGGTG GCAATAATGC GGGGATTGAC ACTATCTGGT ATAATCCTCA 1020

TCACCTCGAA AATCACACAC AAGCCCAGCC GACTTACGAA GTCTATTCTT ACCAAGACTT 1080

GCTGGATTGT TTAGATAAAA ATATTCTTGA AAAGATCACA TTTTAAAGGA GACGAGCTAA 1140

TGACTACAAA AAAGCTAATA TTACTATTGA AGAGTACATT GAAATGTCTG AAGTTGATTT 1200

TAATGAAGCT GTTAATTATG AATTTACATC TGACACTTGT CAATTAGCAA ATAGTATTTA 1260

TCAATCTCTT TTTAAGTTTT TTGATAAGAA AAATTTCTCT GGCGATTTAA TTTTTACTTG 1320

GAAATCTCCA TCATTAGTCA AAGAAGGGGA TTATATTGGG AGAAGGGATT CACAAGTAGA 1380

TAATCTTAGA GTAATAGGAA ATATATTTCC GAATTATCTT ACTAATCGAA AATATAGCCT 1440

CAATATGAAT CGTAATGGCT GTATGGGAGA TTTTCCTCAT GACTTTTTTG ATATATACCT 1500

AGATCATGTA GCAAAATATG CCTACGAACA AAAAGTTAAT AATATTAAAG AGTATTATCC 1560

TTTAAAAAAG GCGATTTTAC ACCAAGAGAA TGCATTGTAT TTTCGATTTT TTTCTAATTT 1620

TGACGACTTT TTAGAAAAAA ATTATTTAAA GACTATATGG CAAGTTTCTA AAGAAACTCC 1680

(2) INFORMATION FOR SEQ ID NO: 271:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 598 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 271:

AGCTCGGTAC GTAGTATnTG TGGTGCATAA ATGAGTGAAA AGAGGATAGA GAGGATGAGG 60

CCGATAAGAA CACCGGTAGC TGCATCGTGA AATACTTGTT TTTTCATAGT TCTAATTTCT 120

CCTTGATGGT TTTTAGATAA CGGCGTGAAG AGTAGGTGAA GCTTTCGTTT TTCAAGAAAA 180

TTTCTACCAG ACCGTTTGGC GTGAgCTTGA GGTGAGAGAT GGAATCGATA TTGATGATTT 240

CTGATTGGGA AATTTGGATA AAATTGGTTG GCAAGAGTTT AAGAACCTGA TAGAGTCGCA 300

AATCAATGCT GTAGGTCTGA CTCGCGGTTT CTGCTAGAAC CTTCCGATTC TCGATATAGA 360

AGCGCTGAAT CTTGCCAATC TCAACTAGAT AGACCTGATC ATCGATTTTT CCTTTGATTT 420

TTTCTCTTTG GTCCAGATTT TCTGCGAACT CGATGACTTT CTGGACTTTT TCGGTTTCTT 480

GAGGTGCTTG GACAATCAGC TTTTCCTCCT CGTAAGTCTC ACTAATCTGT AGTTCTACTT 540

TCATAGTTTT CTCTCCTTTT CAGTTATACA AGGTTGTGAT CACTTCCTGT ATATCCGG 598

(2) INFORMATION FOR SEQ ID NO: 272:

1302

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1099 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

CCAGCAAATC AATAACTGCA ATTGCTATAA AATGGATTCT ATAGAGTTTT TTCATGACAA 60

GACCTCCCTC TTTTATCTAA CTTCATTCTA CTCCAAAAGA ATGGGAGTTA CAACTAAAAT 120

GATAAAAATA GCAGAAGGGA GATTCTCTTA AGTTGGCTAG TATTCTTTAT TTGAGTTTCC 180

TTCTATTATC TAACTTCTTC ATCATTCCAG ACAAATAAAG CTCCGATTGC ATTGAGGATA 240

TAAAAGATGT ATTTACCGAT ATTGGCGAAG TTTCCTTGAA TACCAGCTTT TGTCAGCTGA 300

ACGAAATTGT AAATCAACCA AAAGCCCCAC TGAGTTGTTA GTTTTAATGC ATTCAAAGCA 360

TTGGCAATGA GGGACAGTGC AAAGGCAATA GTTGTTACGT AGGCAAGGAG ATTCATCTTG 420

CCCCCATATC CGATATAGTT GGTCACAAAG GCAAAGAGGA AGGCGATGAT GGAAATGATG 480

ATGGCCGCCA ATTTTACCTG TTTTTGGCTC ATTTGGTTGG GTCTGCCTTC TTGCGAAGCT 540

TCCCACTTCT TTATAGCAAA GGTATAAATG AGGAAGGTGA CGGGATAGGT AATGATGGCC 600

GCCTTATTTC CAAGGATATA ATCAATAGCA CCGGACAAAA TGGTATTAAC AATACCAAAG 660

TAATTTCCCC ATTTGCTTAA TTTCCCCGTG AAACGAGTGG ACAACATGGA AATCCCAACG 720

TTGGTTACGG AAATCAATCC AAAGGGTACA AGAGCTGTCC ATGATCCCCA GTCTACAAAT 780

TTATCGAGGT GTGAGTTGAG GTAACCAGAT GCAATCGCAA TCCCAACGAC CAAAGCAACC 840

CCGAAGAGGT CAAACTATTT AGATGTAGCA AAAATTTTTA GTGATTTTTT CATAGGTTAA 900

ACTACCTTTC TTTTTTTCAA ATATTCTCCC ACCAAATGAA AGTAAAATAA AATGATAGAA 960

ATAAAACCCT GAAAATAAAG GTTCTATAAT ATTTGTAGTG GGTAAATCCA CTATAGATAT 1020

TATGGAGCCT ATTTTATTGT AGAAAAAAAG TCCCATATGA CCTATAATGA AAAGCGACAA 1080

AACAACTCAT TAGAAAGAT 1099

(2) INFORMATION FOR SEQ ID NO: 273:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2723 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 273:

1303

CTGGGATTCA CGTGAAAAGG AAGCCCAGAG AGTAGCCAGG TGTACTGCTA GAACAGTGAG 60

TGAAATTGAA TATTACCATA GAGAGTCAAC CCAGATAGCT CAGGCTTTAG TTGAAAATCA 120

AGCTCGTATC GAGGGAATCT ATAAATACTT TAGCCTTAGC ATGCCAGACT ATTTTTACTG 180

GCAATTAGAG CGGAAAGCTT CGCCTTATAT ATCAGTCTCT CTGTATGAAA ATGTTGATGA 240

CCTCTATGTT CGAAATGATT TTGTAACTGG GGTGGCCATT GCTTTTCAAG ATTACAAGGA 300

AGTCTATGTT TCTACTAAAG ACAAACGTAG GkkAGAAAAA ATCAGGGCTG AGGATTTCAA 360

ACCAGCAGGA AATAGTTTTG CCATTCCAGT GTCAGATCCA GTGTCAGATC AAGACTTAGG 420

AGTGATTTAC ATCTCCTTGG ATCCTGCTGT TTTATACCAT GCCATTGATA ATACTAGAGG 480

TCATACTCCG ATGGCAGTAA CAGTGACCTC ACCTTTTGAT ACGGAGATTT TTCATATGGG 540

TGAGACAGTT GATAAGGAGA GTGAAAATTG GCTAGTTGGC TTAACTTCTC ATGGATATCA 600

GGTTCAGGTG GCAGTTCCTA AAAACTTTGT TTTACAAGGA ACAGTGACTA GCTCTGCTTT 660

GATTGTGGGT TTGAGCCTTC TCTTTATTGT CATTCTTTAT CTGACTTTGA GGCAGACTTT 720

TGCTAATTAC CAAAAGCAGG TAGTGGATTT AGTAGAATCC ATTCAAGTCA TTGCTCAAGG 780

CGAAGAGGGG CGTCGGATTG ACATTTCCGA GAAAGATCAG GAATTACTCC TAATCGCGGA 840

GACGACCAAT GATATGTTGG ATCGATTGGA AAAGAATATC CATGATATTT ACCAGTTAGA 900

GCTTAGTCAA AAAGATGCCA ATATGCGAGC CTTGCAGGCG CAAATCAATC CTCATTTTAT 960

GTATAATACG CTGGAGTTCT TGCGCATGTA TGCAGTTATG CAGAGTCAAG ATGAGTTGGC 1020

AGATATCATT TATGAATTCA GTAGTCTCTT GCGTAACAAT ATTTCCGACG AAAGAGAGAC 1080

CCTCCTCAAA CAGGAATTAG AATTTTGCCG TAAATACAGC TATCTCTGCA TGGTTCGCTA 1140

TCCCAAGTCC ATTGCCTATG GTTTCAAGAT AGATCCAGAG TTAGAGAATA TGAAGATTCC 1200

CAAGTTTACC TTGCAACCGC TGGTAGAAAA CTATTTCGCG CATGGTGTTG ACCACAGGCG 1260

GACAGATAAT GTGATTAGCA TCAAGGCTCT TAAACAGGAT GGTTTTGTGG AAATTTTGGT 1320

GGTCGATAAT GGTAGAGGAA TGTCGGCTGA AAAGTTGGCA AATATCCGAG AAAAATTAAG 1380

TCAGAGATAT TTTGAACACC AAGCCAGCTA CAGTGATCAA AGGCAGTCTA TCGGGATTGT 1440

CAATGTACAC GAGCGTTTTG TGCTCTATTT TGGAGACCGC TATGCCATTA CTATAGAGTC 1500

TGCAGAGCAA GCCGGTGTTC AGTATCGTAT TACAATTCAA GATGAGTAGA AAGGGAGAAA 1560

ATGTATAAAG TATTATTAGT AGATGATGAG TACATGGTGA CAGAAGGTCT GAAGCGTTTG 1620

ATTCCCTTTG ATAAGTGGGA TATGGAGGTC GTCGCAACAG CCAGTCATGC CGATGAAGCT 1680

CTAGAATATG TTCAGGAAAA TCCTGTCGAT GTCATCATTT CCGATGTCAA TATGCCAGAC 1740

AAAACAGGGC TTGATATGAT TCGGGAGATG AAAGAGATCT TACCAGATGC TGCCTATATC 1800

CTGCTCTCAG GTTATCAGGA GTTTGATTAT GTAAAAAGAG CAATGAACCT TAGTGTGGTG 1860

GACTATTTGG TCAAGCCTGT TGATAAGGTA GAGCTGGGAA ATCTGCTGGA GAAGATTGCA 1920

GGTCAGCTCG GCGAGAGAGG GAAGAAAAGT CAGACTCTTA GTCAAGAATT AGACGAGGCT 1980

GGATTTGTTA GTTATTTAGG GGATAAGGAG AATTGGTGGA TAGGTCTATC CAAGGAAAAA 2040

CAAGGTTCCT TCACCATTCC CTACTATGTC TTGGGTCAAG ACTGGCAGAT TTTCATTTCT 2100

GGCCACCCCC TAGATGGTTT AGTCGTTACA CCTTTTGAAG CTCCTTATCA AGAACACTTT 2160

GAACGCTGGA AGCTGAATGC TGAGAAAACC CTCTTTTACG GTTCTGTAAA TCTGCAGCAG 2220

TCTGAGAGTC TCTTTGCCTA TTACGAACCG ATTTATAGGG TTATCATTCA GGGAAATCTC 2280

AATCAAATCG TAGAAGAGTT AAATCTCTTG GAGAAGGTAG TTCTTGAAAA TACACCTCGT 2340

GTTTCGATTA CTAAACAGCT TTTTATCCAG TTTGTCATGG ATGTTTTCCA TTTATTTGAA 2400

CATCTCAAAG CTGATGATAT GACGGACATT GTCAAAACCA TTCATGCTAT TCAATCCTTC 2460

GATGAATTGG TTTCTTATAT CAAGGAAACT CTGATCAGCT TTTTCGGTCA ATACCGTATG 2520

AATGAAAATG TGGTCAGTGT GCTGGAAGTC ATTGGTCGTG ATTACCAAAA AGAGCTTTCC 2580

CTCAAGGATA TCAGTAAGGC CCTCTTTATC AATCCTGTCT ATCTAGGGCA GTTGATTAAG 2640

CGTGAAACCG ATTCGACCTT TGCAGAGTTA CTAAACAAAC AACGTATTAA GGCTGCCCAG 2700

CAGCTCTTGC TTTCAACTAG TGA 2723

(2) INFORMATION FOR SEQ ID NO: 274:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 836 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

CCGCAGTTTT TTTAAACCGT ATATAAGTAT AGCATAGTCA AAAAAAGAAT GCAAGATTTT 60

TGCAAACTTT TTTAAAATTT TTCGTAATTT TTCTTTTAAA GTTCTACTGT CAGGACTTGA 120

CCTTGCTTAA CAACCTGTTC TCCGGCGATA TAAACATCAT CTACATCACT AGATTTAACT 180

GCATAAACCA GGTGAGACAG CATATTTTCC TGAGGTTGGA GATGAATTTT CCCTTGTGGT 240

TGAATGACCA GAAAATCTGC TTGCTTGCCG ACTTCCAGAC TTCCTATCTG ATTTTCCATT 300

CCAAGGACCT TAGCCCCTTC GATTGTCAGT ACCTTGAGAG CTGTTTCGAT TGGAAACTGG 360

CTGGCATCCC CACTTTTCAT CTTCTGAAGA AGAGCTGCAG TCCTTCCTTC CTCAAACATA 420

TCTAGATTGT TATTGGAAGC AACCGAGTCA GTCGCAATTC CGACTGCTAC TCCCGCTTTT 480

TGGAGCTGGA TAATTGGAGC AATTCCTGAT GCCAGTTTGA GGTTACTGAT AGGATTGTGG 540

GCGATAGCnA CTTGAGAAGA TGCCAAGCGT TCAATTTCTC TCTCGTTTAA TTCGACCCCG 600

TGAGCAAATA CGGACGGATG ATCTAAATAA CCCAGTTCTT CAAGAAAAGC AAGGGGGCGT 660

TTGCCGTATC GTTTGAGGAT AATTCCTGAC TCCTCCTTGG TCTCCGCCAC ATGGACATGG 720

AGCGGAATAT TTAGCTCTTT TGCCATTTCC AAACTCGCTT CCAGCAAGTC TCTACTGCAG 780

CTATACGGAG AATGAGGTGC TACCATAACC TTGAAATTTG GATTTTTATA TTTTAA 836

(2) INFORMATION FOR SEQ ID NO: 275:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2335 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

ATTTTATTTC ACTTTTTAGG TGGTCTGGGG CTATTCTTAT ATAGCnTCAA GACCATGGGA 60

GACGGTTTAC AACAAGCTGC TGGAGATCGC CTGGGTTTTT ACATTGACAA ATATACTAGT 120

AATCCTTTGT TTGGAGTTCT GGTTGGTATT GGGATGACTG CTCTAATTCA GTCTAGTTCT 180

GGTGTAACAG TTATCACAGT CGGCCTGGTC AGTGCCGGTC TCTTAACCTT ACGTCAGGCT 240

ATCGGGATTG TCATGGGTGC TAATATTGGG ACAACTGTCA CATCCTTTCT CATCGGTTTT 300

AAATTAGGTA ACTATGCCCT ACCTATGCTC TTTATCGGTG CCGTCTGTCT TTTTTTTACG 360

AAAAATCGGA CAGTCAATAA TATCGGACGC ATCCTCTTTG GTGTCGGTGG TATCTTTTTT 420

GCCCTCAATC TCATGAGCGG CGCAATGGCT CCACTCAAGG ATTTACAGGT CTTTAAGGAC 480

TATATGATTG AGCTAAGTAA GAATCCTGTT TTGGGTGTCT TTGTCGGTAC TGGCTTGACC 540

TTGCTAATTC AAGCTTCTTC GGCTACCATT GGGATTTTAC AAAACCTCTA CGCCGGCAAT 600

CTAATTGATC TACAGGGAGC TTTGCCAGTT CTATTTGGTG ACAATATCGG GACAACCATT 660

ACAGCCATCA TTGCCTCTTT AGGGGCTAAT ATTGCAGCTA AACGGGTAGC AGGAGCTCAT 720

GTTGCCTTCA ACGTTATCGG AACAGTTGTC TGCGTTATTT TTCTAGTTCC TTTTACTGTC 780

CTGATTCATT GGTTTGAAGC TACGCTAAAT CTAGCACCGG AAATGACCAT CGCCTTTGCT 840

CACGGAACCT TTAATATTAC CAACACCATT GTCCAATTTC CATTTATCGG AGCTCTGGCT 900

TACTTTGTAA CCAAGATTAT TCCTGGAGAG GACGAGGTTG TCAAATACGA ACCCTTATAT 960

CTTGATGAAC ATTTCATCAA ACAGGCCCCA TCTATCGCTC TAGGAAATGC TAAGAAAGAG 1020

CTCTTGCACT TAGGAAACTA CGCTGCTAAA GCCTTTGACC TTTCCTATAA GTACATCATT 1080

GACTTGGATG AAAAAGTTGC TGAAAAAGGG CATAAAACCG AAGAAGCAAT TAACACCATC 1140

GATGAGCAAT TAACACGTTA TCTCATTGCC CTTTCAAGCG AAGCTCTCAG CCAAAAAGAA 1200

AGTGAAGTGC TTACCAATAT CCTTGATTCC TCCCGTGATT TGGAACGGAT TGGAGACCAC 1260

ACGGAGGCTC TACTCAATCT GACTGACTAT CTTCAACGGA AAAATGTTGA ATTTTCTGAT 1320

GCCGCCTTGA AAGAATTAGA GGAAGTTTAC CGCCAAACTA GTGACTTTAT CAAAGATGCT 1380

CTGGATAGTG TGGAAAACAA TGATATTGAA AAAGCACGCA GTCTTGTAGA ACGTCATGAA 1440

GCAATCAATA AGATAGAACG TGTTCTCAGA AAAACCCACA TCAAACGCCT CAACAAAGGC 1500

GAATGTTCAA CACAAGCTGG GGTCAACTTT ATCGACATCA TCTCACACTA CACTCGTGTA 1560

TCAGACCACG CTATGAACCT TGCTGAAAAG GTTTTTGCAG AACAAATCTA AGAACCAAGA 1620

AGCTATCCAT CATAATTGGA TGGCTTTTTA CTTTTTCCTA AGCAAGACTA GGATGAATGA 1680

AACTGAAAGA GTATTCTGCA GATATATAGT CCCCAATTAT TCACCCCAAA TCTAAAAACC 1740

ATCCAGAATC CTTGCCTTAG CTTAGATCCT GGATGGTTTC TTTTTTCACC CAATGGGTGT 1800

TTTTTACTAG ACAAAAAAGA GTTTCCCCTT TATGGTATAA GTGTAGAAAA AAACACAAAA 1860

AGAAAGGAAA CTCACATGAA CAGTTTACCA AATCATCACT TCCAAAACAA GTCTTTTTAC 1920

CAACTATCTT TCGATGGAGG TCATTTAACC CAGTATGGTG GTCTTATCTT TTTTCAGGAA 1980

CTTTTTTCCC AGTTGAAACT AAAAGAGCGG ATTTCTAAGT ATTTAGTAAC GAATGACCAA 2040

CGCCGCTACT GTCGTTATTC GGATTCAGAT ATCCTTGTCC AGTTCCTCTT TCAACTGTTA 2100

ACAGGTTATG GAACGGACTA TGCTTGTAAA GAATTGTCAG CTGATGCCTA CTTTCCAAAA 2160

TTGTTGGAAG GAGGGCAGCT TGCTTCACAG CCAACCTTAT CCCGTTTTCT TTCCAGAACT 2220

GACGAGGAAA CAGTCCATAG TTTGCGATGC CTCAACCTTG AATgGkCGAA TTCTTTTTAc 2280

AGTTTCACCA GCTAAACCAA CTCATTGTAG ATATCGATTC TACCCATTTC ACAAC 2335

(2) INFORMATION FOR SEQ ID NO: 276:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 752 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 276:

CGGATTCACT GTTGTTGACT AATCAATAAC ACAGTAGAAA ATCTCACAGC AGTCTATTAG 60

TTGCTTTTCA TACTAGGCAA GTGACTGAGG CTTGTACTTG GGTACAGCAA GGGAGCTTAA 120

GGCCGTAGAA GAGAAAAATA GTAGACTGAA AACCCGCAAG ACTTCATCAT TTCGAGAAGT 180

GACGTGGGAG ATGAAAATCG ATTGAACCAC TTACAAGGAG AATAGAAAAT GGCTAAAAAA 240

AGCAAACAAC TTCGTGCTGC TCTTGAGAAA ATCGACAGCA CAAAAGCATA CAGTGTAGAA 300

GAAGCTGTAG CACTTGCAAA AGAAACTAAC TTTGCAAAAT TTGATGCAAC TGTAGAAGTT 360

GCTTACAACT TGAACATCGA CGTTAAAAAA GCTGACCAAC AAATCCGTGG AGCAATGGTA 420

TTGCCAAACG GTACTGGTAA AACTTCACGT GTTCTTGTTT TCGCACGTGG TGCAAAAGCT 480

GAAGAAGCAA AAGCTGCTGG TGCAGACTTT GTTGGTGAAG ATGACCTTGT TGCTAAAATC 540

AACGACGGTT GGTTGGACTT CGACGTAGtT ATCGCTACAC CTGATATGAT GGCTCTTGTT 600

GGACGTCTTG GACGTGTCCT TGGACCACGT AACTTGATGC CAAACCCTAA AACTGGTACT 660

GTAACAATGG ATGTTGGCAA AGCGGTTGAA GAGTCTAAAG GTGGTAAAAT CACTTACCGT 720

GCTGACCGTG CAGGTAACGT TCAAGCAATC AT 752

(2) INFORMATION FOR SEQ ID NO: 277:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2643 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 277:

GTCAACATTG ATTTCAAGGC TGTTTGCTTT CTATCTCCCC TTTTTCATAA TGTATAATAA 60

AATGAAATAA TAACAGGACG AATTGATCGG GACAGTCAAA TCGATTTCTA ACAATGTTTT 120

AGAAGTAGAG GTGTACTATT CTAGTTTCAA TCTACTATAT TTTCGTACAG GTGCTTCAAC 180

CATTTGAACG ATTTCAAATC CTTCTTTTTG GTAAAGATTC TGAGCTCTTT GATTTGCCTC 240

GAAGACATTT AGAGAAATAC TGTCTATATC TCTATTTTCA AATGCTAAAC TAACAAATTT 300

CCTTAAAGCC TTGCTACCTA AGCCTTGCTC CTGTTTCTGG GGGTTGATAA AAAATCTCCC 360

GATATGAAGA TTGCTGTCTT CTAGCCTGAT TTTCTGGATA AATCCCACAA ACTCTTGTTC 420

ATCAAAGATT GAAAAGACTC CTTCCAAGGC TTGAAGTGTC AGTAGAAAAG GAATCCTTGG 480

TCCCATCCAT TGTTCTTGAA AGGATTTGCC TAGGGAGTTG GACCACTGGC ATACAAATTG 540

AGCGTTTTCT GTGCTCACCT TTTCTTCAAA ACGAATTGTC ATCTTTTCCT CACCACCTTA 600

TCTATGTTTC TCCATTATAC TATTTCTCCC ATTTTTTACG AATAGATAAG TATGATTGAT 660

TTTTATTTTT TTCTCGTCGG GAGCATTCTA GCTTCCTTTC TTGGTTTGGT CATTGACCGT 720

TTTCCAGAGC AATCCATTAT CAGTTCAGCC AGTCACTGCG ATTCCTGTCA GACTCCCTTG 780

CGTCCCTTAG ATTTGATTCC GATTCTCTCA CAGGTCTTCA ATCGCTTTCG CTGTCGCTAC 840

TGCAAAGTTC GCTATCCTGT CTGGTATGCC CTCTTTGAAT TAAGCTTAGG ACTCCTCTTT 900

CTGCTTTACT CTTGGGGATG GCTCTCCTTG GGGCAAGTCG TCCTAATCAC CGCTGGTTTG 960

ACCTTGGGTA TCTACGACTT TCACCATCAG GAATATCCCT TACTGGTCTG GATGACTTTC 1020

CAGCTAATCC TAATAGCTTC CTCTGGCTGG AATCTGGTCA TGGTCTCCTT CCTCATACTT 1080

GGAATTTTGG CTCATTTTAT CGATATCCGC ATGGGTGCAG GGGATTTCCT CTTTTTAGCT 1140

TCTTGTGCTC TCGTCTTTAG CGTAACGGAG TTACTGATCT TGATTCAGTT CGCTTCTGCG 1200

ACGGGTATCC TGGCCTTTCT CCTGCAAAAG AAAAAGGAAA GACTTCCTTT CGTGCCTTTC 1260

CTCTTACTTG CTACTTGTTT GATTATTTTT GGTAAGCTAC TGCTTGTCTG ATAAAATCCA 1320

ATTTCTGCCA TATATCCTTC ATGAAATTAT TTCACAGTTA AATTATAAAT TATTTCTTTT 1380

GTACAAAGGG ATGATGTTAT CAAATCGATC TGTTCTTCTA TCTTCTTGAT ACTGATCAAA 1440

AAATTTCATT TCGACTGAAA ATATTTCGCT TATAAACTGT AAACGAATAC TTTGTTTAGA 1500

CATTATAGTC GCTAGACTGA CTAGATGATT ACTCAAAACG ACGTCCAGAA TACTCTTTAC 1560

TTTGCTTGGT TTTTTAACAA AAATTTGATC ATCCAAGGGT TCAATCATTT TGTAACCTTT 1620

TTGCGCAATT TGACGATAAA AGTAAGAATG TTGCTTTGGA GTCAATAATC CTAACTTAAA 1680

AGCTCGATAC TCTAAAGCCT GTATCGAAAC ATTCAAATCC GACTTCAATA AAATATAACT 1740

ATCAGGATTG CTGACACGCT TGCCAACCCT CTCTTCAAAT TTGACTAAAA ACTCTTCTTT 1800

TGGCAATAAA AAACATGATG CAAAATAATT TGCTTCTTGC TCCAAACGAT CGCCATCTTC 1860

ATTCATATCT TTATATTTAT GTAAAAGAAT ATGTCCTAGC TCATGAGCTA AGTCAAAATT 1920

TCGACGTACA GATGATTTAT TCGTTCCTAA CACAATATAA GGTCTTCCCA ATTTTGACCA 1980

TGCGCTATAA GCATCAGCTT GGCCATTAAT TAATCGTTCC ACGATATAGA TGCCTGAACG 2040

TTCTAATTTA TAAAGCAAAT CATGATTATC TTTTGAAATA CCTAATTTTT CCCTGGCATA 2100

AAGAGCCAAT TCCTCAATGG ATTCTCCCTT ATGATAAGAT TCACTCACTA CATTACTTAG 2160

GTCATGAATT ATAATATTAG GTATAATTAC AAAACTTTCA AAATAATCAA TCAAACTATC 2220

TACCTTATGT AAATACATAG TTTGAATATC TATTGTTTTC CGTGTTGCTA GGTCTGCATT 2280

TCTAAAGGCA ATTACAGAAG AATCAAATCG AATGCTCTCT TCTTCCTGTT CAAAATAAGT 2340

TAAATCAACA TGAAATTGGT TGGCCAAATG CATTTTGGTT GATAATTTAG GTTTCGTTTC 2400

GTTGGACTCA AACTGCCAAA TGGCTTGTTC CGTTAAATTA ATTCTCTGAG CTAATTCTGC 2460

TCTACTTAAA CCATTTAACA GCCGTAATTC TTTCAATACC CGACCATTAA ACATTTACAT 2520

ACTCCTTACT ACTTTTGACC TTCTTGTTTT TCTATTCTTG GAATAATTTC AAAATCTTCT 2580

GTTTCCGATA ATTCTGAAAA ATTAGGAATA TCTTGATATT TAGCTTCTTC GAAATGGTAC 2640

GGG 2643

(2) INFORMATION FOR SEQ ID NO: 278:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 582 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 278:

TGACCAGTGG CAAAATGGCT ATCCAAATGC AGATGTTATT ATCGATGATA TCATCTCAGG 60

GCAAGCCTAC GTAGCCTTGG AAGAGGGAGA ACTGCTAGCC TATGCTGCTG TGACCAAGAG 120

TCCAGAGGAG GCCTATGAAG CTATTTATGA GGGAAACTGG CAAGCTGGAG AGTCAGAGTA 180

TCTAGTCTTT CACCGTATTG CTGTGGCAGC AGATGTGCAG GGAAaAGGAG TTGCTCAAAC 240

CTTCTTAGAG GGCTTGATTG AAGGTTTTGA TTATCTTGAT TTTCGCTCAG ATACGCATGC 300

TGAAAACAAG GTTATGCAAC ATATTTTTGA AAAACTTGGT TTTAAACAAG TCGGTAAGAT 360

GCCAGTAGAT GGCGAACGCT TGGCCTATCA AAAATTAAAG AAATAATGCA AAAGAAGTAT 420

GTAAAAATCC TCTACTCCTC ACCAATTGGT ATTCTATCAC TTGTAGCTGA TGACCATTAT 480

TTGTATGGAA TTTGGGTTCA GGAGCAGAAG CATTTTGAGA GGGGACTAGG AGATGAAACG 540

ATAGAAGAAG TTGTwAGTCA TCCTATTTTA GACCCAGTTA TT 582

(2) INFORMATION FOR SEQ ID NO: 279:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 554 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 279:

CCCAAGCTAC TAAGAGACTA AAACTTGCTA GAGAAGCAAG AGAAAGTGTG AATCTTTTTA 60

ATTTCATGAT GAATTTCCTT TCTGCTACCA ATTTAGAGAA ATTTTCTCTA ACCAGCAATT 120

CCCCTAGTAT AACAAGTTCA AAAAATGGAG TCAATTTATC TGCTCACGGT CCAGCAGGTA 180

1310

GCCCCGTACT TCTGAGATAA AATAGAGAGA CCCTGTAACG AACAGCAAGT CTTGAGCGTC 240

TGCCCTTTCT TCAAAATCGC TGATAAATTC TCGGTAAGAA GAAACTATAT CGTAACCTGT 300

CACATCCCTT TCGTCCAAAG CCCCCTGATA GTCAAAGCCG GTCACCTTGA GTTCCACCTG 360

AGGCAATTTT TCAGTCAGAT AACCCAACAT CCCTTGATAA TCCTTACGTT TCAAGGATCC 420

AAAGAGGATT TGAGGTCGAT AGCCTTCCTG CTCTTTTTCT TTGATAAACT CAGCCAAGCG 480

AGTCAAGGCA GGGAGGTTAT GAGCACCATC CAAATAAATC TGTGGGCGAA TACGCTCCAA 540

GCGAsCAGCC CAAT 554

(2) INFORMATION FOR SEQ ID NO: 280:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 766 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 280:

CCGGTTTTTC AAATGAATTT CTTGGTTGTG GCTAAAAAAT ATGCTACACT ATCAATATGA 60

AAATTTTAAT CCCAACAGCA AAAGAAATGA ACACAGACTT CCCAAGTATC GAGGCAATTC 120

CTTTAAAACC AGAAAGTCAG GCCGTGCTTG ATGCCTTGGC TCTCTATTCT GCCAGTCAAT 180

TGGAGAGTTT CTACAAGGTA TCAGCTGAGA AAGCGGCGGA AGAATTTCAA AATATCCAAG 240

CTTTGAAAAG GCAAACTGCT CAACACTATC CAGCCTTGAA ACTTTTTGAT GGGCTTATGT 300

ACCGCAACAT TAAGAGAGAT AAGCTGACCG AGGCGGAACA AGATTATCTT GAAAATCATG 360

TTTTCATTAC CTCGGCTTTG TACGGTGTTG TTCCAGTCTT GTCACCCATG GCTCCTCACC 420

GTTTGGATTT TTTGATGAAA TTAAAAGTCG CTGGTAAGAC TTTGAAGAGC CATTGGAAGG 480

CAGCCTATGA TGAAACTCTG AAGAAGGAAG AAGTGATTTT CTCTCTCTTG TCATCAGAGT 540

TTGAGACTGT ATTTTCTAAG GAAATCAGAG CAAAGATGGT GACCTTCAAA TTCATGGAGG 600

ATAGAGGCGG TCAGCTGAAG ATTCACTCAA CTATCTCCAA GAAAGCGCGC GGGGCCTTTC 660

TAACAGCTTT AATAGAAAAT CAAGTACAAA CTGTGGGGGa AGCACGTCGC TTGAACTTTG 720

CTGGATTTGT TTACCGAGAA GATTTGTCAC AACCACAGGG GGATGG 766

(2) INFORMATION FOR SEQ ID NO: 281:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 901 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 281:

CCGGCCACGG TTCCATCCAA CTTCACAGGT GTGCACTTGA TTGTGTATGT AATTGTCACT 60

AACGGTAGAA TTTCACCTAT CCCTCCTATC TGCTCGCAGT ACCCGCAGAC TTTCTGAAAG 120

AAGAAGATAA CCTACTTATC CGTTGCTATG ATTATACTAA AGTTTCTACT TTTTTGCAAA 180

TAGATTTTTA AATTTTTGGC TAATTGTCTG AATCAGGGTC GGAAGTTTGA CGACCTTGTC 240

ATTGCCTAGT TTTTCGCGTG CAATTTTGAG AATGGCACCT GAGTCTTTTG AAGCAAAGAG 300

GAATTTTCCT TTGTCTGTAA AGACTTCGAA GTGGCGGCTG ATTTTGCGTC CAGTGACATT 360

GGCTCCAATC TGATTGATAT GGCTCCAAGG AATCTGGATA AATTGTTCGA CATTGACATC 420

TGGGTAAAAT TCCAAAGCCT GATCTCCGAC AAGGAATTTC CCAACTTTCC CAGCGATAGA 480

GAGGTAGGAA GTGCCTGTCG TACTGAGGAG TACTGTTTTG TTAAGTGATT GGGCCATGCT 540

TAGTCTTCCT TACTTTCTCC AAAAAAGGCA TTGTAGAGGG CTTTAATTGC TGCTTTCTCT 600

TGGTCTTTAT TGACAACAAA CATAATAGAA ACTTCACTAG AACCTTGAGA CATCATCTGG 660

ATGTTGATTT TGTTTTCAGA TAGAGCGCGT GTCGCAGTAG CAGTCACTCC GATATGGCTC 720

TTCATTTTTT CACCAACAAT CATAATGATA GAAAGGTCGT GTTCGATTTC TGCATGATCT 780

ACTTTAGCCT TTTGAACCAA CTGACGCAGG ATTTCTTCTT CCTTGATGGG AGTTAGTTGG 840

CGAGAACGGA GAATGATAGA AAGAwCGTCG ATACCTGTTG GCATATGTTC CCAACCGATG 900

T 901

(2) INFORMATION FOR SEQ ID NO: 282:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1765 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 282:

CCCTGTTACG TGGATAATAG GGTAAGACTG CTCAGGATTT CCTAACAAAT CCACCGCTTG 60

CTGCATTCGA CCCAAACCTG ATCGAAAATT CAAACCAATC CGACTATGGA GCCATTCTTC 120

TACTTCAAAC ATACACATCT CCTTGACAAA AGTCCAATCA ATTATCGCAT TAAAGTATGG 180

TTACTAATAA AAACAAGGCC AGGATTTTCG TCCCGACCTC TTACCTGGTT AGCTAATAAC 240

TAGCTACTAT GAATGTGAAT ATGGGCTAAA AACATCCACT GGACGTTCCA ACTCTTCCCC 300

ATTTCTGGGA GTTGGGGTAA AAATGTTCAC TGGACGTTCC AACTCTTCCC CATTTCTGGG 360

AGTTGGGCTG ATACAGTCTC CCAGACTGTA TCACTCCTCC ATAAAGCTGT TGAAGACTTC 420

TTCAATCATG TTCCATTCGT CTTCTGAGTC TTCTGGGATT GGTTGCAATT CGCCTTCTGT 480

TCCATCTTCG TTTTCGATGA ATGAGTAAGC TTGGATTTCA ACTTGTCCGT CTTCGTCTTC 540

TTCTGCGTTA ACTGGTACTA GAAGAACATA GTTTTTACCA AATTCTTCTT TTCCATCAAT 600

TGTCAAAAGG ATTTCAAACA AGGTTTCATT TCCTTGCTCA TCTACTAGTG TGATTAGTTC 660

ACGTTCTTCG TGGTCGTGGT TATGATCGTG TGACATAGCC TCGCCTTTAT ATTAAAATTT 720

TCTATCTAAA TAATTTTGTA AAATCAGCTG AGCTGCTAAC TTATCAATGA CTTTCTTGCG 780

CTTATTGCGA CTGATATCTG CTTGTTCAAT CAACATGCGC TCAGCAGCCA CTGTTGTCAA 840

GCGTTCATCC TGATAGTCTA CTGGTAAACC AAAAAACTCT TCTAGCTTTG CTCCGTAGCT 900

TGACTAGCTT CTACGCGCGG TCCACTTGTA TTGTTCATGT TTTTAGGCAA GCCCACTACA 960

AATCGTTCCA CCTTGTAAGT ATCAACCAAT TCCTTAACGC GGTCAAAACC AAATTGGCCT 1020

TGTTCTTCAT TTATCTGGAT GATTTCAAGC CCTTGAGCTG TAAAACCAAG CGGATCGCTA 1080

ATCGCCACCC CTACCGTTTT TGAACCGACG TCCAATCCCA TAATTCTCAT AGGTTATAGA 1140

TCGACTCCTT GTCCTTTGAG GTAGTAGCGA ACCAATTCCT CAACGATTTC ATCACGCTCA 1200

TACTTACGGA TTTGATTTCG TGCATTATTA TAACGAGGAA CGTAGGCAGG GTCTCCACTC 1260

AATACGTAAC CTACGATTTG GTTAATTGGG TTGTAaCCCT TATCGTTCAA CGAAGCATAA 1320

ACATCTGTCA AAGTTTCGCT AATTTCTTTT TTATTGGAAT CGTCCAATTT AAAACGTACT 1380

GTTTCTTCAG TAAATCCCAT TCTAACACCC TCTTTCCTTA GAATAGTACC ATTATAGCAT 1440

AATTCCTTAC CTTCTACAAT TCAGGCAGTC TATTTATTTG GATTTTCTAT TGTTCTGTCG 1500

CGCCATTTGC CAATCTATCT GAAATATATT TGCTTGGTTC ATTTTTCAAA AGATTTTCCA 1560

AACCAATATT CTTCAGATGT TCCAACTGGG AAGCCTTCTT GACATCCAGA ACTTGAAAAT 1620

CAAAACTAGT CGTTGTTTGA AGTTCCGTTG CGCTCAATAG TTTTGTTTCA AGTTTGAAAC 1680

CTGCCAATTT ACGAGCTTCA ATGATAGACT TATCCTTCTC CTCCGCTTCA AGAAGAGCTT 1740

TTTGAGTTTC CTCCACTCCA TGTTG 1765

(2) INFORMATION FOR SEQ ID NO: 283:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1346 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 283:

CTTATCCATT CACTTTCTTG TCTGTTATTC TATAAATCTT ACTCCTAAGT ATACCACATT 60

TGCCCCTAGA TGTGAACGAG AGAAACGCTC TAGACATTGC CAAGAAGGAA AAAAAAGGGT 120

ACAATGTAAC AAAATCAAGG GAGGTCTGGA ATGAAGAAAC AAAGCAAGTA CAAAGAGGTC 180

GTTTCCTATC TGAAAAATGG TATCGAGTCT GGACGATTTC CGACGGGTAG TCGCCTGCCT 240

TCTATCCGTC AACTGAGCCT TGACTTTCAC TGCAGCAAGG ACACCATTCA ACGAGCCCTG 300

CTGGAATTAC GGCACGAACA ATACCTCTAT GCCAAGCCTC AGAGTGGCTA CTATGTATTA 360

GAACAAGGGC AACATCAAGA CCTAGAAATC GAGGTTACCG ACGAACATGC CAGTGCCTAT 420

GACGATTTCC GACTCTGTGT CAATGAAACC TTGATTGGCC GAGAAAACTA CCTCTTCAAC 480

TACTATGACA ATCAAGAAGG ATTAGAAGAC CTAAGACAGT CCATTCACAA ACTCCTCTTT 540

GAGCAAGCTC TCTACTGCAA GGCTAACCAA CTAGTACTGA CTTCTGGAAC CCAACAAGCC 600

TTGTTTATCC TCTCTCAAAT ATCCTTTCCT AGACAAGCCA AGGAAATCTT GGTGGAACAG 660

CCAACCTACC ATCGGATGAA TCGCCTCTTG ATTGCACAGG GGCTGGACTA TCAAACGATT 720

GAACGAGGCA TTGATGGGAT TGACTTGGAG GAGCTGGAAG GCCACTTCAA AACAGGAAAA 780

ATTAAGTTTT TCTACACCAT TCCCCGATTT CACTATCCCC TGGGACATTC CTATTCTGAG 840

CAAGACAAAC GATCTATTCT TAACTTAGCT GCCAAGTATG ATGTCTATAT CGTAGAGGAC 900

GATTATCTGG GTGATTTGGA CTCCAAGAAG GGCCAAACCT TCCACTATCT TGATACAGAG 960

GAGCGTGTCA TTTATATCAA GTCCTTCTCG ACCAGCCTTT TTCCTGCCCT TCGTATTACA 1020

GCACTCATTC TTCCAAATGC TATCAAAGAA GCATTTGTGG CCTACAAAAA TATCCTAGAC 1080

TACGACAGCA ACCTCATTAT GCAAAAGGCC CTGTCACTCT ATATTGACAG TCAATTGTTT 1140

GAAAAAAATC GTTTGGCTCG CTTGACCAAT CATGAATCTT ACCAAAAACA AATCGAGGAA 1200

AGGATAACTA AAACACCTTG TCCCCTTCCT CATTATTCCC TACACGATGG yTTATTGCTA 1260

GACCTGAGAC AGTATCCTAA AATCGCCAGT CTCAAACACA GTCAACTGGG cTTGGACTTC 1320

TTTGAAGAGG CCTATTTAAG CACCTG 1346

(2) INFORMATION FOR SEQ ID NO: 284:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 900 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 284:

CTATATTCAG AATATGCCAA AAATTCGGAA TGGTATAAAT TTGCGGAGGG TTCATTTGAC 60

ATATTTAGAA AACTCCCCCA AAGAATTAAT TTTAAGAAAG ATTTTTCTAG AATTTTGGCC 120

CCCTTTATTA TTAATTTGCT TAAATTAATC AATAATTATC TAGAGAATAA AGAATACGAG 180

TGGATTGACA AGAATGGAAA TATTTTTTCC TCTCTAGTAT TTTATTTAGA AGATTTAATC 240

TATCCTTGGA TTGTTAAACC TTTGGTTTTA GAGATAAATT CATTGCGTGA AAAAGGTTTA 300

CTTGAAGGGG AATCGGAGCA GCAACGGTAC AAATATTTTA TAACATTGTT TGACAAGGAA 360

GAGAATATAT TAAATTTTTA TAACAAATAT CCCGTTTTAC TGAGGCAAAT ATCGGAGTCT 420

TGTCTTCGGT TCTATACTTA TTTTATAGAA ATTTTATCAA ATTTAGAAAA TGATTTTAGT 480

GTGCTAGAAG AAGAATTAGG GCTAAGGGGG AAATTAAATG ATATAAAATT TGGAAAGGGT 540

GATACACACA GCCAAGGAAA AACTGTTTTG ATACTCTTCT TTGATGACGC GAAAATTGTT 600

TACAAGCCTA AAAATTTAAT AATCAATAAC TCACTAAATA CTATTGCTGA GTATATCCGA 660

AAGGTTGATG AAAAAATTAG GATAAGAATA CCTCGAACTA TTGCTTATTC GGATCACAGC 720

TATGAAGAAT TTATTGATTA TCTACCTCTA GAGCAAAAGA AAAATTTACC TGAATATTAT 780

TATAATTTTG GTGTGCTTTT AGCATTTATA TATTTATTTA ATGGGAGTGA TATACATTTT 840

GAAAATTTAA TTTCCTATGG AGATATGCCT GTAATAATAG ACTTTGAAAC AATGTTACGG 900

(2) INFORMATION FOR SEQ ID NO: 285:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 862 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 285:

TTATTTAGCA GAGGCAGTTT TAAATGTGAA GGATTTGGTC AGTCAAACAG TTTTTTATCA 60

GCAGATTATT GGTTTAGAAA TCCTATCTCA AACGGATACA GAGGTCGTTC TGGGACTTGG 120

AGGAAAAGCC TTGGTACACT TGATTCAAGC ACAAGAGGGT GGAGAAGTAA GGGAACATTA 180

TGGTCTTTAC CATCTGGCTA TTCTTTTGCC GACACGAAAG GCTTTGGCGG ATGTCTTGAA 240

GCACCTGACG GATTTACAGA TTCCTCTTGT TGGCGGTGCA GATCACGGTT ACAGTGAGGC 300

CCTTTACTTA GAGGACTTGG AGGGAAATGG CATTGAACTC TATCGAGATA AGCCAGTTTC 360

CACATGGGAT ATTCGAGAAG ATGGACGTAT TATCGGGGTG ACTGAAGTCC TTGCGGCTCA 420

GGATATCTAT GAGTTGGGGG AAAGAGTAGA GCCTTTTATC CTAGCAGAGG GTACGAGAAT 480

GGGGCATATT CATCTTTCTG TCAAGGATAG TCGAAAGTCC AGACAGTTTT ATCAAACGGT 540

GTTAGGGCTC GAGGATAAAT TCAGTGTGCC TAGTGCTAGT TGGATCGCAG CTGGGGACTA 600

CCATCATCAT TTAGCAGTCA ACGAATGGGG AGGAAAAGGT CTGGATCCGC GTAAACAAGT 660

CCTACCAGGT TTAGCCTACT ATGTCATCGA AGTCGCACAT AAAGAAGAAC TGTTAACGAT 720

TGCCCAACGA GCACAAGAAG TTGACGCACC AATCAAATGG ATGACATCGA TCCAATTGGA 780

AATCACAGAC TCAGATGGCA TCGTGACCCG TATTCGTTTA GCTAGATAGA TGGTATGTGA 840

TGAAGGTAGA GCATCAATTG TA 862

(2) INFORMATION FOR SEQ ID NO: 286:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 650 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 286:

TCGTTTACAA GATCGCTAAA ATGCATCTCA TGATCGCGAC CACGAATTCC AAGATAGCAC 60

GCGCTACCTC AATCATAGAT AGTTCACTTT TTTCTTGCCC AGCAAATACT TCTAATTCCA 120

AAGCGTTTCT CCTCATTTAT ACTACTATCG CCAGAGCGAA CAGACTCTGA CCTCATTTTA 180

TCATTTACTC TTTATTTTAC GATAATTTTG CGGAATAGTC AAAGGTTAAG GGGGAGAAAG 240

TGGCAGGATT AGACTAATTC CAATATAAAA CTCATTCCTT TTTCTGTTGC TCCATTTTCC 300

ACAAATCCAA GCGACTTGAA ACACCTCCTA GAAGCATGAT TGTAGGTGTA GATTTTCTTG 360

ACTCTCAATT CTTTCCATCC TTTTACTCGA GCCAATTCAA TCAAAGCACT TAGAATCTTT 420

TTTCCAAGTC CTCGATGTTG GTAAGCGGAA TTCCCAATCA CAATGGGGAG ATTATCCTGA 480

GATAGTGTAA TATCCCCAAT TGGAAACCAT TCTCCCTTCT CCTTGACTTC AATCCAAAAA 540

AGCTCACCAT GCCGATyCAr ATAGGAATAC ATGGCTTCCA AGGTCGcTtG ACTGTAAGGA 600

AGCTTCACCC CATCTACGAG GtAAcCAAGT TCACATCCGT GATACCAAGC 650

(2) INFORMATION FOR SEQ ID NO: 287:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1119 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 287:

GATAGCAATC CGCTTCAGAA ACTTCTCGCT TACCTCTAAC TCCGATCGCT AGTTTGGGAG 60

AAGATACTTC CATTCTCATA CTATCTGTTG GCTTTGCAGG CTGTAAAAAC AACTTTTCTC 120

TTGCTACTTC CTGAAAATCT GAATCTTGCA GTTCTTTGCT TTCAAAATAG TCCTGTACTC 180

GCTCCACATC AAAATTCCCA GCTAAAGACA GAGACATGTT TACAGGTTTG TAAAACTTTG 240

TAAAATTTTC TTGCAAATTA GTTAGATTGA TTTGGGAAAT GGACTCCTCA CTTCCAACTA 300

TATCAGTTGC TAAAGGTGTA CCAGGATACA AATTCGCTAA AGTTGAAAAG AATAAACACG 360

AATCTGGATC ATCTTGGTAC ATTTCTCGTT CTTGCTGAAT AATATCCTGC TCTGTCAGAA 420

TGGAAGCTTC AGTAAAGTGT GCTGATGTTA CCAATTCATC AAGTAAATCT AAATTTTCTA 480

AAAAATAATC CGTTGCTGAA AAAAGATAGT TTGTTTTTGT AAAGCTTGTA AAGGCATTAC 540

TATCTGCACC TAGACTCGTA AAAGCCGACA TCAAATCACT AGAATCTTCT CTCTCAAATA 600

ATTTATGTTC AAGAAAATGA GCAATTCCTC CAGGATATTG TTTTACATCT CCGTCAACTT 660

CTGTGACAAA CGTATCTACC GAACCAAACT GTACAGTGAC ACTCCCGTAA ACCTCTTTAA 720

ATTCCTTTTT AGGCAAAAGA GCAACTGTCA ATCCGTTGGC CAAACGAGTT CGATAAACCA 780

TTTCTTTTAC AGCTGGATAG TATTTTTCTT CAAAAACAAC CTTTGTCATT CTATTCCTTC 840

CATAAAGTAA ATCGCTTGTA GTTTCACATT ATTAGCTACT CTACAAATAG CATCTTTGTC 900

AATTTGTTCA AGCTTTGCAA TCCAACTTTT AAAGTCTGCT GAAGATTTTC CAAATAAGGC 960

ATTTTGATAA GCACGTTCAA TCAATGAAGA ATGATTATCT TGAGAAAGTA ACAACGACCA 1020

ACGAATCATT TCCTTGGTCT GATTTAACTC AAACTCTGTA AAAAAACCTT TTTTTAAATC 1080

AAGCCGTTGA TTATTCATCA ATTTACGAGC CTGGTTACG 1119

(2) INFORMATION FOR SEQ ID NO: 288:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 540 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 288:

ACGCCCTCGC GGGGACATGA CGAATTCCCC GTTCATCACG AAGGCCGCCG AGGAGTGGGG 60

GGTGCCGTCC AAGTCAAAAG CGGCCCCACA TCGATTCAGT TCCCCGACGA ACAGCCCTTT 120

CCCCCAGCGT TCCTGGCTTT GCAACCGTTT CACAACAGCC TCGTAAAGTA GGCCGGACAA 180

GGCAGACGGA CTCCAAAGGA GTTCTTCCAT CTGCAAGTGC GCCTGCGTTA TGTGATCCCG 240

GTCTTTTGCA TGTGTGTGGC ATGAATGCTG TTCCCAATCC CACTCCAGAA CATTCTCCTC 300

AAAAGTGCGC AACGTCGCCC TGAATGAATC CTGCCTTGTA GTCGTGACCA TTCCTATGAA 360

GGGTCGCAGA GGATTTTCCC CGAGTGCAAG CGCATCCTCC GGCTCAAATC GGGTGCATTT 420

CACAGTCCCG CTCAACGCTA GCCCGATCCC TTTTTGGCAT GGTGACTCAA GCGTCCTTTC 480

AAACAAAAGC TCCTCATCCG CTCCAACCGG CCCGACGTAG ACGCGTAGAC CGAAGTCGTC 540

(2) INFORMATION FOR SEQ ID NO: 289:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1949 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 289:

AAAGAATTCG ACCAATTCAA GGTTGAGGCA TCGCAAACTA TGGACTGTTC CCCCGTCAGT 60

TCTGGACAGA AAACGGGATA AGGTTGGCTG TGAAGCAAGC TGCCCTCCTA CCAACAATTT 120

TGGAAAGTAG GCATCAGCTG ACAATTCTTT ACAAGCATAG TCCGTTCCAT AACCTGTTAA 180

CAGTTGAAAG AGGAACTGGA CAAGGATATC TGAATCCGAA TAACGACAGT AGCGGCGTTG 240

GTCATTCGTT ACTAAATACT TAGAAATCCG CTCTTTTAGT TTCAACTGGG AAAAAAGTTC 300

CTGAAAAAAG ATAAGACCAC CATACTGGGT TAAATGACCT CCATCGAAAG ATAGTTGGTA 360

AAAAGACTTG TTTTGGAAGT GATGATTTGG TAAACTGTTC ATGTGAGTTT CCTTTCTTTT 420

TGTGTTTTTT TCTACACTTA TACCATAAAG GGGAAACTCT TTTTTGTCTA GTAAAAAACA 480

CCCATTGGGT GAAAAAAGAA ACCATCCAGG ATCTAAGCTA AGGCAAGGAT TCTGGATGGT 540

TTTTAGATTT GGGGTGAATA ATTGGGGTTT TACAATATCA ACTCCCATGA TAGTCATGAG 600

ATGACTCTTC ACGAATTGAC GTGATGACTG TCCTTCCTTT TGCATAATTA CCTCCGAAAC 660

ACAAAAAAAG GGGTAGACAA TCTAGTGTCT ACCCCCGAAA GTTTATTAAA ACAAAAATCC 720

TGCCAAAGAA TTTTTGGCAG GAAACCAAAT CAATTTATCA GTTTCTATCA ATCGCTTATC 780

GCTCTCAAAG ACTGGTAAAT AGGGATTCCG CAATCAAATT GCGATACTCT ATTATTTAAG 840

AGTAACTGAA GCTCCAGCTT CTTCCAATTT AGCTTTGATT TCTTCAGCTT CTGCAGTTGC 900

AACGCCTTCT TTAACAAGTG CTGGTGCACC GTCAACAAGT TCTTTAGCTT CTTTAAGACC 960

AAGACCAGTG ATTTCACGTA CAACTTTGAT AACGCCAACT TTTTTGTCGC CTGCAGATGT 1020

CAATTCAACG TCGAATGAAT CTTTAGCAGC ACCAGCATCA GCTGCATCAG CTGCAGCAAC 1080

AGCTACAGGA GCAGCTGCAG TTACACCAAA TTCTTCTTCG ATAGCTTTTA CAAGGTCGTT 1140

CAATTCAAGG ATTGAAGCTT CTTTAATTTC AGCAATAATG TTTTCAATGT TCAATGCCAT 1200

TGTTATTTCC TCCAAATAAG TTTTAAATTT TATAATAGTT TTTTTCGTAG CTAGksTACG 1260

CTGTGTAGCT TAAGATTAAG CCGCGTCTTC TTTGCTTTCT GCAACCGCTT TGACTGCAAG 1320

AGCAACGTTG CGCACTGGCG CTTGAAGTAC AGAAAGGAGC ATAGAAAGAA GTCCTTCGCG 1380

GTTTGGAAGA GTTGCAAGTG CAAGAATCTC TTCTTTAGAT GCGACAGCGC CTTCGATTGC 1440

ACCACCTTTA ATTTCAAGTG CTTCAGCGTT TTTAGAAAAG TCGTTCAAGA TTTTCGCTGG 1500

TGCGATAACA TCTTCATTAG AAAATGCTAC TGCAGATGGT CCAACAAATA CAGATGCAAG 1560

ATCTTCAAGA CCAGCTTTTT CAGCTGCACG ACGCAAGATT GAGTTTTTAA TAACTTTATA 1620

CTCAACTTCG CTTCCACGAA GCTCACGACG AAGAACTGTA TCTTGCTCAA CTGTCAAACC 1680

ACGAGCGTCT ACAACGACGA TAGATGCAGC AGCTTTCATT TTTTCAGCTA tACGTCAACT 1740

AGTTCCGCTT TTTTAGCAAT AATTGCTTCA CTCATTAGTG TGTTCACCTC CGTAATTATT 1800

TTGCTTGGGG AATTTTTCAA AAAGAAAAAC GCGCCCAATC CTAGACACGA AAGTACAATA 1860

CGCTTCTTTT TACATGATAC GTTTTGTCCT CGGTAGGATA TTTATGAGTC GAGCTCCCCT 1920

ACTGTCTTAG GCAGTTTTTT TAGATACGG 1949

(2) INFORMATION FOR SEQ ID NO: 290:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1023 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 290:

GGACTGTTTG ATCTTATACA GTAGCTGCTT GATCCAAGCT TTCACCGATA GCGGCTAGGC 60

GCTCGATAAC TTCAGCTTGT GTCAATTCAT TTTTTGAAAC ATAGCGGTTA CGTGGGTGAA 120

CACGGCACTC GTGTGAGCAT CCACGAAGGT ACTTGTCTTC ATTTTCTTCT GATGTCAAGA 180

TACGACGGTT ACAGAATGGA TTTCCACAGT TGACATAACG TTCACATGGT GTTCCATCAA 240

ACCAGTCTTT CCCTACGATA GTTGGGTTGA CATGGTTGAC ATCAACGGCA ATACGCTCGT 300

CAAAGACGTA CATTTTCCCA TCCCAAAGCT CACCTTGAAC TTCTGGGTCT TTACCGTAAG 360

TTGCGATTCC TCCGTGCAAT TGGCCGACAT CTTTGTAGCC TTCACGGACC ATCCAGCCTG 420

AGAATTTCTC ACAGCGAACG CCACCTGTAC AGTAAACCAC GACACGCTTG TCCATGAATT 480

TTTCCTTGTT ATCACGGACC CATTGTGGTA ACTCACGGAA GTTGCGAATA TCTGGGCGAA 540

TAGCTCCACG GAAATGTCCT AGGTCGTACT CATAATCGTT ACGTGTGTCA AGGACAACGG 600

TATCTTTATC AAGAAGCGCT TCTTTGAACT CTTTTGGAGA CAAGTAAGCA CCTGTTGTTT 660

CAAGTGGGTT GATGTCATTG TCAAAGTCGT TGTCTTCCAA ACCAAGGTGG ACAATTTCTT 720

TCTTGTAGCG AACAAACATC TTCTTGAAGG CTTGTTCATT TTCTTCGTCA ATCTTGAACC 780

AGAGTTCTTC CATTCCTGGA AGGCTGTGAA CGTAgTCCAT GTATTTTTGA GTTGTTTCAT 840

AGTCACCTGA AACTGTTCCG TTAATTCCCT CGTCAGCGAC TAGGATACGG CCTTTAAGGn 900

CGATTGATTT ACAGAAAGCC AAGTGGTCTG CAGCAAATTG CTCTGCATTT TCAATTGGAG 960

TATAAAGGTA GTAAAGTAAG ACACGAATAT CTTTTGkCaw AAGATTTGTA TCTCTTTATC 1020

TAT 1023

(2) INFORMATION FOR SEQ ID NO: 291:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3831 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 291:

ACTATGAACA AGACCCAGAA AAAGTAGCCT TATTTCTTAA GAATTTTAAT AGTTTAAAGC 60

ACCTAGCACC TGTTTAGATT GACGAAACAG GATTCGATAC TTATTTTTAT CGAGAATATG 120

GTCGCTCATT AAAAGGTCAA TTAATAAGAG GCAAAGTATC TGGAAGAAGA TATCAGAGGA 180

TTTCTTTGGT TGCAGGTCTA ACAAAATGGTG AATTAATCGC TCCAATGACT TACGAAGAGA 240

CGATGACGAG CGACTTTTTT GAAGCTTGGT TTCAGAATTT TCTCTTACCA ACATTAAACA 300

CACCATCGGT TATTATTATG GATAATGTAA GATTCCATAG AATGGGGAAG CTAGAACTTT 360

TATGCGAAGA GTTTGGGCAT AAACTTTTAC CTCTTCCTCC CTACTCGCCT GAGTACAATC 420

CTATTGAGAA AACATGGGCT CATATCAAAA AGCACCTCAA AAAGGTATTA CCAAGTTGCA 480

ATACCTTTTA CGAGGCTTTT TTATCCTGCT CTTGTTTCAA TTGACTATAT TAGAGGCGAG 540

ACATTTTTCG GTTCTTTGTC AACTGTAGTG GGTTGAAGAA AGCGAAGATC TAGAAAGGAC 600

AAATTTCGTC CTTTCTTTTT TGAAGTTTTC AAAGTTCCTA AAACCAAAGG CATTGTGCTT 660

GATAAGTTTG ATGAGATTAT TGGTGGCTTC CAGTTTGGCG TTGGAATAAG GTAATTGAAG 720

GGCGTTGACG ATTTTCTCTT TATCTTTGAG GAAGGTTTTA AACAAAGTCT GAAACAGAGG 780

TGGAAAAGCA AGAGCTGATA GAGATTATAG TGGTGTTTAA AGTCTTCGGA ATAGCTCAAA 840

1320

AGTTTATCTA GAATTTCTTT ATTAGTCAAG TGCATACGAA AAGTAGGGCG ATAAAATCGT 900

TTATCACTCA GTTTCTGACT ATCTTGTTGA ATGAGCTTCC AGTAGCGCTT GATAGCCTTG 960

TATTCATGGG ATTTCGGATG ATGGCTTGTG TTCTGCTCTC AAGAACAGTT ATGATATTGA 1020

GTTTATCAAA GTCCTGAGCA ATAAAGCTCA TCTCCATCTC CCGATTGAAA CAGTCACTCC 1080

CCGGACTGTT TCAACsTCCT AGGACATAAT CTCAGGAAGA eGeGAAAAAT CATGCTCAAA 1140

GTGAAAATCA TTGTTCTTGC GAATGACAGT TGAAGTTGAA ATAGACAACT GATGATCAAT 1200

GTCGGTCATA GAAGTCTTTT TAATTAGCTT CTGAGCAATC TTTTGGTTGA TGATACAAGG 1260

AATTTGATGA TTCTTCTTGA CGATAGAAGT CTCAGCGAGC TCCATTTTTG AGCAATGATA 1320

GCACTTAAAA CGGCCTTTTC TAAGAAGAAT TCTAGTTTGA ATTTTTTTAT ACTAGAAAAT 1380

CAGAACCATA ATACCTATAT AAAAATATTA TAGTTCTAAT AGGATTTACC CAAAAGTTTT 1440

AAGGCGGTCT TTTTAGAACT TTAATTGTTT GAAATTTAGG TAGCAAATTT GTTTCTATTT 1500

TGTCAACTTT TCCTATTTTT ATCTTGTTGA GGCTGGTATT TTAACAATTC AGGAATTGAT 1560

AGTGAATGTG TAAAATTTTT TGTTAGAATA AGTTTATAAA AAAGAAAAGG AGTATTTGAT 1620

TATGTTACAA AAAATTTATG AGCAGATGGC TAATTTCTAT GATAGTATTG AAGAAGAGTA 1680

TGGTCCTACA TTTGGTGATA ATTTTGACTG GGAACATGTT CATTTTAAAT TTTTAATTTA 1740

TTATTTAGTG AGATATGGCA TTGGTTGTCG TAAGGATTTT ATTGTTTACC ATTATCGTGT 1800

TGCTTATCGT TTGTATCTTG AAAAATTGGT AATGAATCGG GGTTTTATTT CTTGTTGAGG 1860

TAATTTTAGT AAATTTCCGA ACTAATTTAC TCTTTTATGG AAAGATGATA GTAAATAGCT 1920

AGTAATTTTT CTAAATCATT TTTTAATAGT TGGAAATAGC AAATCTTTCT ATTGTTTCTT 1980

CTTGATAAAA AGGCGATTTT TTATTATAAT AAATTGTAAG ATATAATTGC AGGTGAGAGT 2040

CCTGCCATGT ATGTGAGAAA GGAAGAGCCT GATGGCTCAG ACAAGATTAT GACTTCAGTT 2100

GTTGTTGTAG GTACCCAATG GGGTGATGAA GGTAAAGGGA AGATTACAGA CTTCCTTTCA 2160

GCGAATGCAG AAGTGATTGC ACGTTACCAA GGTGGTGATA ATGCTGGTCA CACGATTGTG 2220

ATTGACGGTA AGAAATTTAA GTTGCACTTG ATTCCATCTG GGATTTTCTT CCCTGAAAAA 2280

ATATCTGTCA TTGGGAATGG TATGGTTGTA AATCCTAAAT CTCTTGTAAA AGAGTTGAGC 2340

TATCTTCATG AGGAAGGTGT AACAACTGAT AACTTGCGTA TTTCTGATCG TGCGCATGTT 2400

ATTTTGCCTT ATCATATCGA GTTGGATCGC TTGCAAGAAG AAGCTAAGGG CGACAATAAG 2460

ATTGGTACGA CAATTAAGGG AATTGGTCCA GCTTATATGG ACAAGGCTGC TCGTGTTGGA 2520

ATTCGTATTG CAGATCTTTT AGATAAAGAT ATTTTCCGTG AGCGTTTAGA ACGTAACCTT 2580

GCTGAAAAGA ATCGTCTTTT TGAAAAATTG TATGACAGTA AAGCGATTGT TTTCGATGAT 2640

ATTTTTGAAG AATATTACGA ATATGGTCAA CAAATCAAGA AATACGTGAT AGATACATCT 2700

GTTATCTTGA ATGATGCGCT TGATAATGGC AAACGTGTGC TTTTTGAAGG TGCACAAGGT 2760

GTTATGCTAG ATATCGACCA AGGTACTTAT CCATTTGTTA CGTCATCAAA CCCTGTAGCT 2820

GGTGGTGTGA CAATTGGTTC TGGTGTCGGT CCAAGCAAGA TTGACAAGGT TGTAGGTGTA 2880

TGTAAAGCTT ATACGAGTCG TGTAGGAGAT GGTCCTTTCC CAACTGAGTT GTTTGATGAA 2940

GTGGGAGAAC GTATCCGTGA AGTGGGTCAT GAATATGGTA CAACAACTGG TCGTCCACGT 3000

CGTGTAGGTT GGTTTGACTC AGTTGTGATG CGTCATAGCC GTCGTGTTTC TGGTATTACT 3060

AACCTTTCTT TGAACTCTAT TGATGTTTTG AGCGGTTTGG ATACTGTGAA AATCTGTGTG 3120

GCCTATGATC TTGACGGTCA ACGTATTGAC TACTATCCAG CTAGTCTTGA ACAATTGAAA 3180

CGTTGCAAGC CTATCTATGA AGAGTTGCCA GGTTGGTCAG AAGATATTAC CGGAGTTCGC 3240

AATTTGGAAG ATCTTCCTGA GAATGCGCGT AACTATGTTC GTCGTGTGAG TGAATTGGTT 3300

GGCGTTCGTA TTTCTACTTT CTCAGTAGGT CCTGGTCGTG AACAAACAAA TATTTTAGAA 3360

AGTGTTTGGT CCTAAGAGAT TTTTAAGATT TGTTTAAGAT AGGTCGGGTA TACTATAGAC 3420

GGTTACAAGA AGACCTCCTA ACTTGTTGTA ACAAATATCC TAAACTTTTC TTTTTCATAA 3480

TAATCTCCCT ATAGAGTCAC CGCATTCGGT GGCTTTTTTT GTGTTGGGAT TCATGATATA 3540

ATAATAAAAT CGATAAGTAG GAAAAGAGAA AAGAGATGTA TTATACGCTT GAAGAAAAAG 3600

AAGTCTTTAT GAGGGAGGCT TTGAGAGAGG CTGAGATTGC TCTTGAACAC GATGAAATTC 3660

CAATTGGTTG TGTGATTGTC AAAGATGGGG AAATCATTGG TCGTGGGCAT AATGCGCGTG 3720

AGGAATTACA GCGAGCGGTT ATGCATGCGG AAATTATGGC TATAGAGGAT GCGAACTTGA 3780

GTGAGGAGAG TGCGCTTGCT GGATTGCACA CTTTTTGTGA CCATTGAACC G 3831

(2) INFORMATION FOR SEQ ID NO: 292:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1441 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 292:

CCGCTGTTCC AACCGCAACA TACCATAGTC CGTACGGGAT TCGAACCCGT GTTACCGCCG 60

TGAAAAGGCG GATGACTTAA CCCCTTGACC AACGGACCTG AGTTGTTATT TTCAACTCTT 120

ACTATTATAC AGTCTTTTCA AACTTTGTCA ACTACTTTTT CTAATTTTTG TTTATTTTTT 180

1322

CAACTTATAG TAAAAAAAGC CAGAATTATA CTGACTCTTC TATCGCTCAT TAAACTTAGA 240

AGCACGTTCT TTTCCCCACC AATAAGGGAT TAGTTCTGCG ACTTTAACTG TTTTTCTTAT 300

ATTATAGTCC ATCATGAATT CTGCATCTTT ATTTTCAGCA TTAAGCTCTA AAAGGAATTC 360

TCTACAAGCA CCGCAAGGCA TGGCTGAACT TCCACCATAA GGTGGTTTGT CTCGAAAGGC 420

TAATACTTTC TTAACCTTAG TTTGTCCTGA AAATTGGTAC ATATTGAAGA GGGCCGCCCG 480

TTCTGCGCAG AGATGGAAAA CACCACAGGT TCCCTCCATA CAGAATCCTG TAAATATTTG 540

TCCATCTCCT GCTTCTACTG CAGCTACAAC ATGATTGGCA TAAACAAAGT CTGATACTTC 600

ATGTGGATTG TATAGTTTCT GTGCTTCTTC GTACATCTTT TCCCAGATGT CCATTATTGT 660

ATCCTCTTTA TTTAGAGATT TCTTTTAGCA TGTTTTCGAT ATGCTGAATT GATTTTTCAC 720

GTCCAAGCAA GAAAATTGTA TCTGGTAATT CTGGCCCATG CATTTCGCCT GAAACTGCGA 780

TACGAATAGG CATGAAAAGA TTTTTCCCTT TAATACCTGT TTCTTTTTGG ACTGCTTTAA 840

TTTGTGGGAA GATATTTTCT GTCACAAATT CATCATCTGT CATCGCTTCA AGTTTTGCTT 900

TGAATGCTTC AAGAACTGTT GGAACTGTTT CACCCGTCAT GACTTCGCGC TCTGCTTCTG 960

TCAATTCTGG GAAATCTGAG AAGAAAAGAT CTGTCAATGG GATAATCTCA TCTACTGATT 1020

TCATTTGTGG TTTATAGAGC TCAACTAATT TTTCAGCCTT GTCAGTCAAA CGGCCTGCTT 1080

CCTCTAAGAA TGGTTTTGCC ATTTCAAAGA TGGTTTCAAG GTCTGCATTC TTGATATAAT 1140

CATTGCTCAT CCAGTCTAGT TTTTTCTGAT CAAAGGCTGC TGGTGACTTG CTGAGGCGGT 1200

TTTCATCAAA AAGTTTAATG AATTCTTCAC GAGAGAAAAT CTCATCCCCA CCACCTGGGT 1260

TCCAACCAAG AAGAGCAATA AAGTTAAAGA CTGCTTCTGG AAGGTAACCT TTCTTTCGGT 1320

AATCTTCGAT AAATTGAAGT GTATTAGTAT CACGTTTAGA TAACTTCTTA CCAGTTTCAG 1380

AGTTGATAAT CAAGTGTCAT GTGACCGAAC TCTGGAGCTT CCTCAACCTA AGAGCGGGTA 1440

T 1441

(2) INFORMATION FOR SEQ ID NO: 293:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4398 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 293:

CGGCTTATGT AGTGGCAATC TTTCTACGTA AGCGAAACGA GGGGAGATTA GAGGCGCTAG 60

AAGAAAAAAA AGAAGAACTA TACAATCTTC CAGTAAATGA TGAAGTAGAA GCTGTAAAAA 120

ATATGCACTT GATTGGACAA AGTCAAGTGG CTTTCCGTGA ATGGAATCAA AAATGGGTCG 180

ATTTATCTCT CAACTCTTTT GCCGATATTG AAAATAATCT CTTTGAAGCA GAAGGCTATA 240

ACCATTCATT TCGTTTTCTC AAGGCCAGTC ATCAAATTGA CCAAATTGAG AGTCAAATTA 300

CTTTGATTGA AGAAGATATT GCGGCAATTC GCAATGCTTT GGCAGACTTA GAGAAGCAAG 360

AATCTAAAAA TAGTGGTCGT GTTCTTCATG CTTTGGATTT ATTTGAGGAA CTTCAGCATA 420

GAGTTGCTGA AAATTCAGAA CAGTATGGTC AAGCCTTGGA TGAAATTGAA AAACAATTAG 480

AAAATATCCA ATCTGAATTT TCACAATTTG TAACCTTGAA TTCATCGGGT GACCCTGTGG 540

AAGCCGCAGT GATTTTGGAT AATACAGAAA ATCACATTTT GGCCTTAAGT CATATTGTGG 600

ATCGTGTTCC AGCCTTGGTT ACGACGCTTT CTACAGAATT GCCAGATCAA TTACAGGATT 660

TGGAAGCCGG TTATCGTAAA CTAATTGATG CTAATTATCA TTTTGTTGAA ACGGATATTG 720

AAGCGCGTTT CCACTTGCTT TATGAAGCAT TCAAGAAAAA CCAAGAGAAT ATTCGTCAGT 780

TGGAATTGGA TAATGCCGAA TATGAGAATG GACAGGCACA AGAGGAAATC AATGCCTTGT 840

ATGATATTTT TACTCGAGAA ATTGCTGCTC AGAAAGTAGT GGAAAATCTA CTTGCAACTC 900

TTCCAACTTA TCTTCAACAT ATGAAAGAGA ATAATACTTT ATTGGGAGAA GATATTGCAC 960

GTTTGAACAA GACCTATTTA CTTCCTGAGA CAGCTGCAAG CCATGTTCGT CGTATTCAGA 1020

CAGAATTAGA GAGTTTTGAG GCAGCTATTG TTGAGGTAAC TTCAAATCAA GAAGAACCAA 1080

CCCAAGCTTA TTCAGTTCTT GAAGAAAATC TTGAGGATTT ACAAACTCAA CTAAAAGATA 1140

TTGAAGATGA GCAAATTTCA GTTAGTGAGC GCCTGACACA AATTGAGAAA GATGATATTA 1200

ATGCACGTCA AAAGGCCAAT GTTTATGTCA ATCGTCTCCA TACTATCAAG CGATACATGG 1260

AAAAACGCAA TCTGCCAGGT ATTCCACAAA CTTTCTTGAA GTTATTCTTT ACGGCAAGCA 1320

ATAATACCGA GGATTTAATG GTTGAGTTAG AACAAAAAAT GATTAACATT GAATCTGTTA 1380

CCCGAGTTCT TGAAATTGCA ACGAATGATA TGGAAGCTTT AGAAACGGAA ACTTATAATA 1440

TTGTACAATA TGCAACTTTG ACAGAGCAAC TCTTGCAATA TTCTAACCGC TATCGCTCAT 1500

TTGATGAACG CATTCAAGAA GCATTTAACG AaGCTTTAGA TATTTTTGAA AAAGAATTTG 1560

ATTATCACGC TTCATTTGAC AAGATTTCTC AAGCATTGGA AGTGGCAGAG CCTGGTGTAA 1620

CCAATCGCTT TGTTACCTCA TATGAGAAAA CACGTGAAAC GATTCGTTTT TAATAAAAGA 1680

AAAAGATTTT ATTGTGTGAG GAGCAGAATC AAATCTTTTT CTATAGTTGT GGGGAGATTT 1740

ACTTCATTTT CTCCTGAGAT TGAGTTTTTG CCCAGCCGAT TTATCCACTA CCTCAAAACA 1800

GTGTTTTATA CTCTTCGAAA ATCTTTTCAA ATCACGTCAG CGTCGCCTTA CCGTACTCAA 1860

GTACAGCCTG AGGCTAGCTT CTTAGTTTGC TTTTTGATTT TCATTTAGTA TTAAAGTGAT 1920

TTCGCCAGTC TTATCTGCAG CTTCAAATCT GTACTTTGAG TAACTTGGTA ACCGTCCAAT 1980

AACGAAGTCT ATTGAAAAAT CTCCAGACTA GAGAACTCAC GGATAGTTCC TAATCTGGAG 2040

ATTTCTTATT TGCACTTTTC TTGTACAACT TTAGTCCACG GTAAATAGAC CTCTAAAACC 2100

TCTTTGTTTA CGAGAGTTTC CTCGTTTGGA AGACATTCTA GAAGATAGGA TAGATATTTC 2160

TCGCTATTTA TACTAGACTA AAATCAAAAA GCATTATATA ATAGTGATAT GAAATCAACT 2220

AAAGAAGAAA TCCAAACCAT CAAAACACTT TTAAAAGACT CTCGTACAGC TAAATATCAT 2280

AAACGCCTTC AAATCGTTCT ATAGTAAAAT GAAATAAGAA CAGTACAAAT CGATCAGGAC 2340

AGTCAAATTG ATTTCTAACA ATGTTTTAGA AGTAGAGGTG TACTATTCTA GTTTCAATCT 2400

ATTATATTTC GTCTGATGGG CAAATCTTAT AAAGAGATTA TAGAACTTTT ATAGTAGATT 2460

GAAATAAGAT GTGAACAACT CTATCAGGAA AGTCAAATTA ATTTATAGAA ATATTTTAGC 2520

AGCCAAGGTG TACTGTTATA GATTCAATAC ACTATAGACT GTAATCAAAC AACGATTTGG 2580

CGAAATGTAA AAAAATATGA GGAGTTCGGA CTCGACTCTC TCCTTCAAGA AACACGTGGT 2640

GGTCGTAACC ATGCATATAT GACAGTTGAG GAAAAGAAAG TCTTTCTTGC CCGCCATTTG 2700

AAGGCTGCAG AGGCAGGAGA ATTTGTTACA ATTGATGCCT TATTTCAGGC TTATAAAAAG 2760

GAGTTAGGTC GTTCCTACAC ACGTGATGCC TTCTATCAAC TGTTGAAGTG CCATGGTTGG 2820

CGAAATATTA TGCCACGTCC AGAACATCCT AAGAAAGCAG ACGCTCAAAC CATTGTCGCG 2880

TCTAAAAATA AAATCTCAAT TCAAGAAGAA AAGAAAGCGC TTTAAAACCA GTAGACGTTT 2940

TCGTAAGGTT CGCTTGATGT ACCAAGATGA GGCTGGTTTC GGTAGAATCA GTAAACTGGG 3000

ATCTTGTTGG GCTCCAATAG GAGTAGGTCC ACATATCCAT AGTCACTATA TACGAGAATT 3060

TCGCTATTGT TATGGAGCTG TTGATGCCCA TACAGGCGAA TCATTTTTCT TAATAGCTGG 3120

TAGATGTAAT ACTGAGTGGA TGAACGCCTT TTTAGAAGAG CTTTCACAAG CTTATCCAGA 3180

TGATTATCTT TTACTCGTTA TGGACAATGC TATATGGCAT AAATCAAGTA CCTTAAAGAT 3240

TCCGACTAAT ATTGGTTTTA CCTTTATTCC TCCATACACA CCAGAGATGA ACCCCATTGA 3300

ACAAGTGTGG AAAGAGATTC GTAAACGTGG ATTTAAGAAT AAAGCCTTTC AAACTTTGGA 3360

AGATGTCATG AATCAACTCC AAGATGTTAT ACAAGGATTG GAGAAGGAGG TGATAAAGTC 3420

CATCGTTAAT CGGAGATGGA CTAGAATGCT TTTTGAAAAC AGATGAGTAT AAAAAGAAAG 3480

TCCTCATTTC AATAGAAATC ACGACTTTCT GATGGATTTA TAGTAAAATG AAATAAGAAC 3540

AGGACAAATC GATCAGGACA GTCAAATCGA TTTCTAACAA TGTTTTAGAA GCAGAGGTGT 3600

ACTATTCTAG TTTCAATCTA CTATATTTTT GGAGTGATAG AAAAGCCCTT CATAAGCTAG 3660

TCTACTTGTT CAGGTGCGAG AGCTTTGACA TCTTTTTCTG TACTTAGCCA AGTCAGTTTT 3720

CCGTTCTCAA AGCGTTTATA TAGTAGCCAA AATCCTTGAC CATCCCAGTA AAGGGCTTTA 3780

AAGCGGTCTT TACGTCCACC ACAAAAGAGA AAGACTTGAC CGGAGAAAGA ATCCAATTCA 3840

AAGTGGGTTT TAACTACATA GGCTAATGAG TCTATTCCCT GCCTCATATC TGTCTTGCCA 3900

CAAACAAGGT GAACTTGACC TAAATCACTT AGTTGAATTA TCATAGTACA ATACCTTTCC 3960

TCCGATAATT ATTTTTTATC TAGTATACTG GAAGTTGGGG AATTAGGATA GATACCTTGT 4020

TATGACGCGC TTACGTAACT TGTAACTAGC TGCCTAGTTT GATCTTTGCT TCTTCATTGA 4080

TTAGCAGTAG ATTTCAAAAT GATAAAAACG CATAGTATCA GGTATTGAAA TGTACTGCCC 4140

CAAAAGTTAG ACAGAAAAAA TCTAACTTTT GGGGTGTTTT TGTTATGAAA TTAAGTTATG 4200

ATGATAAAGT TCAGATCTAT GAACTTAGAA AACAAGGATA TAGCTTAGAG AAGCTTTCAA 4260

ATAAATTTGG GATAAATAAT TCTAATCTTA GGTATATGAT TAAATTGATT GATCGTTACG 4320

GAATAGAGTT CGTCAAAAAA GGAAAAAATC GTTACTATTT TCCTGATTTA AAACAAGAAA 4380

TGATTAATAA AGTCTTAC 4398

(2) INFORMATION FOR SEQ ID NO: 294:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 718 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 294:

AGATTTTTAG ACTTTGTCTT TAATCGTTTC TTTTTAGGGA TGATTGCGAC ACCTTCTTTT 60

GGCTATTAAC TTTAGCAGGA GGGATTATCC TTGGTCTAGC GCCGGCTAGT GCCACCTTGA 120

TGAGCTTATA TGCAGAACAT GGTTATAGCT TTCGGGAATA CAGTTTGAAG GAGGCTTGGT 180

CTCTTTACAA GCAAAATTTT GTCTCAAGCA ACCTGATTTT CTATAGCTTT TTAGGTGTGG 240

GTCTAGTTTT GACCTATGGT TTGTATCTCT TGGTGCAATT GCCTCATCAG ACCATTGTTC 300

ATTTGATTGC GACCCTTTTG AATGTCCTAG TAGTTGCCCT GATCTTTTTG GCTTATACAG 360

TATCTTTAAA ATTACAAGTT TATTTTGCCT TGTCCTATCG AAATAGTCTC AAATTATCCT 420

TGATTGGCAT CTTTATGAGT CTAGCAGCTG TGGCTAAGGT TCTCCTTGGG ACTGTGCTAC 480

TTGTAGCAAT TGGTTATTAT ATGCCTGCCC TGCTATTTTT TGTAGGAATT GGGATGTGGC 540

ATTTCTTTAT CAGTGATATG TTGGAACCTG TCTATGAAAT CATCCATGAA AAATTGGCGT 600

CAAAATAGAA TGAAGCAGTT TTGGCTACAT ACGCTTCTAA GAACCTATAG TTCAGTGATG 660

ATCATTATCA TTGCGAGTTT TGCAATCTTA CTCTCTTACG CTGTCTGGGA TTCACGTG 718

(2) INFORMATION FOR SEQ ID NO: 295:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 718 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 295:

TCGGTACCAA AATTCTGGAT TTATACTAGC AAAGATCCAA GAGCAAATTA TTTAACAGAT 60

TTAGGTCTAG TTTTCCCTGA ATCATTAAAA GAATTTGAGA GTGAAGATAG TTTTGCAAAG 120

GAAATTTCTG CAGAAGAAGC AAATAAGATA AATGATGCTG ATGTAATCAT AACTTATGGT 180

GATGATAAAA CTCTTGAAGC TTTACAAAAA GATCCTCTTT TAGGTAAAAT AAATGCAATT 240

AAAAATGGTG CCGTTGCTGT AATTCCAGAT AATACACCGT TAGCAGCCTC ATGCACTCCA 300

ACACCACTTT CAATAAACTA TACTATTGAA GAATACCTAA ATCTTTTAGG AAATGCATGC 360

AAAAATGCGA AATAAAAAAC AAATAAACCT AGGCATAATT TTTATAATCT GCCTAGGTCT 420

TCTTATTACA ATATTTTTGT CATTAAAGCT TGGAACAAAA GAAATTAATA TCAGAGATTT 480

TTTAGCAGCT TTTGGAATGG GTAATACAAA TGATGATTTT ATTAAATCAA TTATATATAA 540

rAGAATACCT AGAACTATTT TTGCAATTTT AGCAGGTTCT AGTCTTGCCA TAAGCGGTGT 600

ATTGATGCAA TCAGTTACTA GAAACCCAAT AGCTGATCCA GGTATACTCG GTATAAACAC 660

AGGAGCAAGT CTTAGTGTAG TAATTGGTCC TTCtTTTTAG GGAATTCATC AAGCATAA 718

(2) INFORMATION FOR SEQ ID NO: 296:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1436 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 296:

GAACTAATCA TTTTTACAGG ATGAGATTTA CAGCAGAGAG TTTGAAGGCT TTATCAAAGG 60

TTTTTCTTGG CATAATGACT TTTCCTCGTT TCCACTTAAT TTTGTGTCTA CTTTATTATA 120

CCAAGTCCAC sCTTAAGTTA GATAATAAAT CTAACTTAAG GAAGCTAGAA GGATGAGAAT 180

CCAGGTGGTC AAGAGTCCCA AACTTAAGCT GATGGGGACA CCCAGAATAA TTTGCTTTTT 240

GAAGGCAAGG CCACGTTCCT CTATATTGGG AAGTGAGAGT TGAATGAGAG AACCAGCTGA 300

TGAAAAGGGT GAGATATTAG TAGATAGAGC GCCAATAACG GTGGCTGTTG TGAGTAAGTG 360

AATATCAATC TGAGGATTTT GAGCACTGAT GATAGCAATG ATGGGAAAGA GGGCTGGAGC 420

TACAACGGAT AGGGTGGAAC TAAAGAGTGA CATCACTCCG GCTATCACAC AAAAGAACAG 480

AGGTAACCAG AAATGAGGAA TGGTTGTTGT CATGAGGTGC CCTATCAGTG TGACTAAACC 540

TGACTTGACC GCTAGAGACA TTAGTAAGCT CATGCCGCAG AGCATGATAA TTGTAGCCCA 600

GGGAACCTTA GCTAAAATGG CTTCTTGCTT CCCTAATTTG AGCCTTAAGG CGAGGCAGAC 660

CATGAGTATT GAGACAAAGC CAATATCAAA TGTTTTTTGA TAAGTAGCTA TCCAGGCGAT 720

GTTTGGGAAA ATGAGATGCA ACAAGGGAAA AAGCCAAACC AAAACCATGC TGCTGATCAT 780

GAGCAAGGTG GTTTGTCTTT GAACCTTGCT GAGGAGTGGT GGTTGGTCAA TAGTCAAGGA 840

TGAGTTTGTT CTTCCCTTAC TATAGTGACT GTAACAGGAT AATAAAAGCA AGACGATGAG 900

TGGGTAGATA ATGCTGACGA TAAAGATATG ATTGCCAAGT GAAAAAGCTT GCTCTTCCCA 960

TCCCATTTGC TTAAACAGGC CTTGAAAGAC AATGCCTGAG CTACTGGTTA TCAAATTAGC 1020

CCCTCCTGAA GCTCCCCAAT TGACGGCTTG AGCTCCAATC AAAGGGTGTT TGTCCGCTTT 1080

TTGACAGAGG GTAATCGCTA GAGGACAGCA AACGGCCATA GTAGTGAAAA ATCCAGCACC 1140

TAAAGCAGAC AAAAGGGTTG CCATCAGGTA TAAAATCATG TAGAGGGCGT TAGGGTGGGT 1200

GCGTGTGCGG TAGAGAATGT GTTGAGCCAA AACATCAAGA GTACCGTTAG TTGTTGCAAC 1260

GTTATAAAAG AGAGAGACGC TAAAAATGGT AAAAAAGAGT GAGGTTGGCC AAAAATGAAG 1320

AAGTTCTTTG GGGCTTAATC CCATGAGAGT GGTTGCGATG AGGTAAGAAA AAGCAATAGC 1380

CAGCAGGCCA ATATTGATTT TGGTGCGGTA ACCAATTCCA ATGGCTAGAG CAATGG 1436

(2) INFORMATION FOR SEQ ID NO: 297:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1696 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 297:

CCATTTGGGA AAGAACGTAA GAGTTTGCAG GGTGAGATTC CAGAAGAATT TTCAATGTCA 60

GCCGTTGACA TGTCTATGAT TGACCACATT CCAGATATGA TTGAAAATGG TGTGGACAGT 120

CTAAAAATCG AAGGACGTAT GAAGTCTATT CACTACGTAT CAACAGTAAC CAACTGCTAC 180

AAGGCGGCTG TGGATGCCTA TCTTGAAAGT CCTGAAAAGT TTGAAGCTAT CAAACAAGAC 240

TTGGTGGACG AGATGTGGAA GGTTGCCCAA CGTGAACTGG CTACAGGATT TTACTATGGT 300

ACACCATCTG AAAATGAGCA GTTGTTTGGT GCTCGCCGTA AAATTCCTGA GTACAAGTTT 360

GTCGCTGAAG TGGTTTCTTA TGATGATGCG GCACAAACAG CAACAATTCG TCAACGAAAT 420

GTCATTAACG AAGGGGACCA AGTTGAGTTT TATGGTCCAG GTTTCCGTCA TTTTGAAACC 480

TATATTGAAG ATTTGCATGA TGCCAAAGGC AATAAAATCG ACCGCGCTCC AAATCCAATG 540

GAACTATTGA CTATTAAGGT GCCTCAACCC GTTCAATCAG GAGATATGGT TCGTGCATTA 600

AAAGAAGGAC TCATCAATCT TTATAAGGAA GATGGAACCA GCGTCACAGT TCGAGCTTAA 660

GAAAGGAAAA GGAAATGATA GAGGCACAGG GTTTCTTAGT GGATAAGCAA ACAAGATGCA 720

TTCATTACCA TAGCAAGCTG GATATTATTG CTTTACAATG CTATGATTGT AAAAAGTATT 780

ATGCTTGTTA TCGGTGTCAT GATTCATTAG AACATCACCC TTTTGAGCCG TATCCCTTAT 840

CTTTGATACA GGATAAGCCT ATTTTATGTG GTGTTTGTCT AAAACTACTA ACATATAAGC 900

AATATAAAGA AAGCTTAAGT TGCCCCTTTT GTTTTTCTCG CTTTAATCCA GGTTGCCAAA 960

ATCATAAGGA ACGCTATTTT AAATAGCAAA TCATCTAGTT TTGAAGTAGG AGAAAACTCA 1020

ATTTCAAGAG AAAATGAAGT AAATCTTCCC ACAATAAAAC GCATAATATC AAGATTGTTC 1080

AATACCTGAT ACTATGCGTT TTTAAGATTT TAAAGACTTT TTTCCTTTAT CTGGTATTTT 1140

GACTACTTGT TAAAACTGGG TTAATTTTCG ACTGTTTAAT AGTTATTATG CAAAGTCTAA 1200

AAGGTTAGAA TTGTCAAAAC AATCCGTCTA GAGTATGCGT GATGCCAACC GTGGTGGATG 1260

TTCTCAGTCA TGCCGTTGGA AGTACGACCT TTACGATATG CCATTTGGGA AAGAACGTAA 1320

GAGTTTGCAG GGTGAGATTC CAGAAGAATT TTCAATGTCA GCCGTTGATA TGTCTATGAT 1380

TGACCATATC TCAGATATGA TTGAAAATGG TGTGGACAGT CTAAAAATCG AAGGACGTAT 1440

GGAGTCTATT CACTATGTAT CAACAGTAAC CAACTGCTAC AAGGCGGCTG TGGATGCCTA 1500

TCTTGAAAGT CCTGAAAAGT TTGAAGCTAT CAAACAAGAC TTGGTGGACG AGATGTGGAA 1560

GGTTGCCCAA CGTGAACTGG CTACAGGATT TTACTATGGT ACACCATCTG AAAATGAGCA 1620

GTTGTTTGGT GCTCGTCGTA AAATCCCTGA GTACAAGTTT GTCGCTGAAG TGGTTTCTTA 1680

TGATGATGCG GCGGTA 1696

(2) INFORMATION FOR SEQ ID NO: 298:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1022 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 298:

CCGAGTTTAT TATGGTTTCT TCGGAATTTA TCTCAAAGAT TGAATTTGCT TGCAATAAGA 60

AAGAAAGTCT TTATAGTCAA AGCAAATTTA AGTATGCGAT TCGTTCGATG TTCGCAGGTG 120

CATTTTTAAC CTTCAGTACT GCTGCAGGTG CAGTTGGGGC TGACTTGATT AATAAAATTG 180

CACCAGGTAG TGGACGCTTC CTCTTTCCAT TCGTTTTTGC TTGGGGCTTG GCCTACATTG 240

TTTTTTTGAA TGCCGAGTTG GTCACTTCAA ACATGATGTT CTTGACTGCT GGTAGTTTCT 300

TAAAAAAAAT CTCTTGGAGA AAAACAGCTG AGATTTTACT ATACTGTACC TTGTTCAACC 360

TTATCGGAGC CTTGATAGCA GGGTGGGGCT TTGCTCATTC GGCAGCCTAT GCGAATCTGA 420

CACACGATAG TTTCATCTCA GGTGTTGTTG AGATGAAGTT AGGCCGCTCA AATGAATTGG 480

TCTTGCTTGA GGCGATTTTG GCAAATATTT TTGTAAATAT TGCGATTCTG TCATTTATTT 540

TGGTCAAAGA TGGTGGTGCC AAACTTTGGC TTGTGTTGTC AGCTATTTAC ATGTTTGTAT 600

TCTTAACAAA CGAGCACATT GCGGCGAACT TTGCTTCTTT CGCGATTGTG AAATTCAGTG 660

TTGCTGCGGA TTCAATTGCC AACTTCGGTG TTGGAAATAT GCTTCGCCAC TGGGGTGTGA 720

CTTTCATCGG AAACTTTATC GGAGGAGGCC TCTTGATGGG TCTTCCATAT GCCTTCCTCA 780

ATAAAAACGA AGATACTTAT GTAGATTAAG AAAATGAGCA CGATTGAGTC GTGCTTTTTT 840

CATTTTCAAA ATAAGGTAAT AGCTATTTCT TATATCAAAA TATAGAAAAC TGATATTTGT 900

ArACTATAAC TCAAGGTGCT ACAATATCCT TAATAAAATA ATATGGAGGT CACCTTATGA 960

CTTGTGATTT TAAATnTGAA ACTCTACAAC TACATGCTGG TCAAGTTGTG GCTCCAGCTA 1020

CT 1022

(2) INFORMATION FOR SEQ ID NO: 299:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 663 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 299:

CCTTAAGTAA TCTCTGATAA TATTTTCTTT ATTAGCATAG GGGAATATCG ATATAATGGC 60

TTCATTATGA GTGGCAGGAA TATCCAATAT GGCAACTTTT CCAATAGATA ATTTAAAACT 120

CATTAATAAA GTTCCTTTAG GTGAAATGTC TATTTTCTTT GATTTTAATG CTAATTTAGA 180

AATAGATTCT CTCGCATTAG TTACATAACC AGATATAGGC ATATCTGATA TAGATACCCA 240

AGGTATTTCA GTTCCCCAAA AAGTAGCTTC ACTGCGTGGA GGAGTTTTTC CTATTCTGAA 300

GTTAACTAGG CTAGCAAATT TAATATATCT CCATGCTTCT GGGATTTCAT ATATAGGATA 360

AGAGGTTGTT TCGTCTTTGT TCCCATAATA AGAGTTATCA TCTCCTTGGG AAACAATAGA 420

AATGTCCAAA TCTTTCTTTT TAATCTTGCC TTCTTCAAAG AGTTTTTGTT TTTCTGCTCG 480

TATTTTTTCA AGTAAAACTT CGACTGATTC ATCATTTGGG TCTTGTTCAA CTAATTTTCC 540

TTGCATAGCA TATTGAAGAA TAGATTTTTT TAGTTTATCT GGAAATTCTT TATCTAGCTG 600

TTCTAGTCTA TTATAACTTT CAGCATATTC ATCTACTTTT TCTAAAGCTG ATTCGATTGC 660

TTC 663

(2) INFORMATION FOR SEQ ID NO: 300:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 881 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 300:

CGTCGCTGAA CATGTCAACA GCAAATTAAA CTAAACAAAC TAAAATTATG TGATACTTCA 60

CATAATTTTC TTTAGAAAAT ATTATCAGAA GAAAGTTGAG AAAAATGGCA GAAAAAACAT 120

ATCCTATGAC CCTTGAGGAA AAGGAAAAAC TTGAAAAAGA ATTAGAAGAA TTGAAATTGG 180

TTCGTCGACC AGAAGTGGTA GAACGCATTA AGATTGCCCG TTCATACGGT GACCTTTCAG 240

AAAACAGTGA GTACGAAGCA GCTAAGGATG AACAAGCCTT TGTCGAAGGA CAAATCTCTA 300

GCTTAGAAAC AAAAATCCGC TATGCTGAAA TCGTCAATAG CGACGCAGTT GCCCAGGACG 360

AAGTAGCGAT TGGTAAAACA GTCACCATCC AAGAAATTGG TGAGGACGAA GAAGAAGTTT 420

ATATTATCGT AGGTTCAGCT GGTGCAGATG CCTTTGTAGG TAAGGTTTCA AATGAAAGCC 480

CAATTGGGCA GGCCTTGATT GGCAAGAAAA CAGGTGATAC AGCAACCATT GAAACGCCTG 540

TTGGTAGCTA TGATGTAAAA ATCTTGAAGG TTGAAAAAAC AGCCTAAAAA CAGAAAAAGG 600

AGTGGGGAGG CGATGTGCTT CACTCACTCC TTTTTCCATT TTGCTACTCT TCGAAAATCT 660

CTTCAAACCA CGTCAGCGTC GCCTTGCCGT ATGTATGGTT ACTGACTTTG TCAGTTTCAT 720

CTACAACCTC AAAACAGTGT TTTGAGCTAA CTTCGTCAGT TTCATCTACA ACCTCAAAAC 780

TATGTTTTGA GCTGACTTCG TCAGTTTCAT CTACAACCTC AAAACCATGT TTTGAGCCGA 840

CTTCGTCAGT TTCATCTACA ACCTCAAAAC TATGTTTTGA G 881

(2) INFORMATION FOR SEQ ID NO: 301:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 949 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 301:

CCTTTTTTAA TACAAGTTAT TTTGATTTAA CCGGCTTGTC TTGAGCTGTC TGCAAAGCTG 60

TGGCAATCGT ATCTGCATAC AATTTTGCTC CTGCTTCGAT AGTGCTACTC TCACTCCCGA 120

AATGAACCTG GTCTGTTCCA GCCCAAATTT CTGGATGCTC TTTCGCAACT TGATTCCAAT 180

CTGCTATCGT AATGTAAGGT GTCTTCTCTG CCAATTCTCT CATATAGGCA GCAGCCTTCT 240

CAACGATGGC ATAGGTCTCT TTTGTCTTAT CTCCCTCATA AGGAGTCACC AAAATCATAT 300

GGTGTCCCTT AGGAAGATTT TTCACGATAC TGTCCCAGTC ATCCTTGTAA TTCTCAGGAT 360

TATTTACCCC AGTCGCAATG ACCACCGTCT TAGGTAAAAA TTTATTCTGG CTATTATTTA 420

GCATGATTTC ATTTGCGGTC TTGGTTGTTA CGCTGACCTG CGCGTTAATC TGTGCTCCAG 480

GAAGAGCTGT CTGTAGTGCT GTATTTGCCC TTAAAGCCAC TGAGTCACCA ATTAACATAG 540

TGCCATCAGC AATTCCCAAA CTGTTTGCAT CTGCCCGTTC TGCCATCACC TTGGTCTGGC 600

CAATATTTGT TGCAGCTTGC TTCAAGCCAT TGACAGTCAA GTCTGTCTCA AACGCTCCCA 660

CTTGTGGTGC CAACAAGGTC ACCGTGCAGA CAATGATGGT CAAGATTCCT GTACCTGCTG 720

CAAGAATTGC GTGAATATAA GGCAGGGGAC GAAsGGTTTG GACAATAGGT GTGTTCTTGC 780

CTGCAATCCA AGGTTCCAAT ACATAAAATG ACAGACTGGC AAAGCCATAA GAACAAATCA 840

GAGTCAGTAA TACAGCAAGA AGATTTGATG TCAACTGTGA GAAAATGATA TAGAAAGGCC 900

AATGGAAAAG ATAAACCGCA TAGCTAGTAT CCGCTAAAAA GCTGATAAT 949

(2) INFORMATION FOR SEQ ID NO: 302:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 622 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 302:

AAGATATATT TTTTACACAG AAGTATGCAA AAGTAAAGAG TGCAAAAAAT GGAATTAAAG 60

CGAAAATAAA AGCCGTGTAC AGGCGACCAA ACCAACGTAC ACGGCTAAGG AAAAATAACA 120

AAACTCAAGC AAAGGCAAGG CGCGTGGTTT TGTTAGGTAT TTAGCAAGGG GACAAACCCC 180

TTTGTAAATA ATCTCCTCTT ATTTTATCAA AATTAGAGGA AAATGACAAC TTAATTTATA 240

AAAAGGAAAA ATGGAGGATA TAAATGGAAA TTCTGTCTAA AGAAATACAG TTACAGGGCT 300

TACAACTTCT TAAACAGACT CTTGAAACTT TAGTTGAGCT AGAAAAACAA CGATCTAGTA 360

AGTTAGATTT AATTTCTCGT AAAGAATTAA TGGATCTGCT AGGTATAAGT GCTACAACCC 420

TTGATAACTG GGAGGATCTT GGTCTTAAAC GATATCAGAC TCCGATGGAT GGAGCTAAGA 480

AAGTATTCTA TCGTCCGTCA GATGTGTATT TATTTTTAGC AATAAAATAG GAGTTATGAA 540

ATGAAAATTG TTACTTTCAA ACCAACTAAA CAAATAGACG ATGGGTTTTA ACTGCCAGGT 600

ATTGACATTC TATTTGTCTC AG 622

(2) INFORMATION FOR SEQ ID NO: 303:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1929 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 303:

CGCTAACTTG CAAACAAAAG AAGAACGCAA ACTCCACAAA TCCTTTACGC AGAAACTCAA 60

TCTCATCTAC TTACCTTGCT GACTTGGTAG AGTATGTTGC AGACAAAGAC TTCTCAGTAA 120

ACGTAATTTC TAAATCAGGT ACAACAACTG AACCAGCGAT TGCTTTCCGT GTCTTTAAAG 180

AACTCTTGGT TAAGAAATAC GGTCAAGAAG AAGCTAACAA ACGTATCTAT GCAACAACTG 240

ACCGCCAAAA GGGTGCTGTT AAGGTTGAAG CAGACGCTAA CGGTTGGGGA ACATTTGTTG 300

TTCCAGATGA TATCGGTGGA CGCTTCTCAG TATTGACAGC CGTTGGTTTG CTTTCAATCG 360

CAGCATCAGG AGCTGACATA AAAGCTCTTA TGGAAGGTGC GAATGCAGCT CGCAAAGACT 420

ACACTTCAGA CAAAATCTCT GAAAACGAAG CTTACCAATA CGCAGCTGTT CGTAACATCC 480

TTTATCGTAA AGGCTATGCA ACTGAGATCT TGGTAAACTA TGAGCCATCA CTTCAATACT 540

TCTCAGAATG GTGGAAACAA TTGGCTGGTG AATCAGAAGG AAAAGACCAA AAAGGTATCT 600

ACCCAACTTC AGCCAACTTC TCAACTGACT TGCACTCACT TGGTCAATTT ATCCAAGAAG 660

GAACTCGTAT CATGTTTGAA ACAGTTGTCC GTGTTGACAA ACCTCGTAAA AACGTGCTTA 720

TTCCTACTTT GGAAGAAGAC CTTGACGGAC TTGGTTACCT TCAAGGAAAA GACGTTGACT 780

TTGTAAACAA AAAAGCAACT GACGGTGTTC TTCTTGCCCA CACAGATGGT GATGTACCAA 840

ACATGTATGT GACTCTTCCA GAGCAAGACG CTTTCACTCT TGGTTACACT ATCTACTTCT 900

TCGAATTGGC AATTGCCCTT TCAGGTTACT TGAATGCTAT CAACCCATTT GACCAACCAG 960

GTGTTGAAGC TTATAAACGT AACATGTTTG CCCTTCTTGG AAAACCAGGA TTTGAAGAAT 1020

TGAGCAAAGA ACTTAACGCA CGTCTATAAT AGAAGAAAAG AGTGGTTTGC CCACTCTTTT 1080

TACTCTCTTT ATCCATAGAA ATTGGACTCA GCCAAGACTT GTGATATAAT ATAGAAAGCA 1140

AAAAGGCAGA CGCCTAGATA ATAGGAGAAA CTATGTCAAA AGATATCCGC GTACGTTACG 1200

CACCAAGTCC AACAGGACTA CTACACATCG GAAATGCTCG TACAGCATTG TTTAATTACT 1260

TGTATGCGCG CCATCATGGT GGAACATTTC TCATCCGTAT CGAAGATACT GACCGTAAAC 1320

GCCATGTCGA GGATGGTGAA CGTTCACAAC TTGAAAACCT TCGCTGGTTA GGCATGGATT 1380

GGGATGAAAG TCCAGAATCA CATGAGAATT ATCGCCAGTC TGAGCGTTTG GACTTGTATC 1440

AAAAATATAT TGACCAACTA TTAGCTGAAG GAAAAGCCTA TAAATCTTAC GTTACAGAAG 1500

AAGAGTTGGC AGCTGAACGC GAACGCCAAG AAGTAGCTGG CGAAACACCA CGCTACATCA 1560

ATGAATACCT TGGTATGAGT GAAGAAGAAA AAGCAGCTTA CATCGCAGAA CGTGAAGCAG 1620

CAGGGATCAT CCCAACTGTT CGTTTGGCTG TCAATGAGTC AGGTATCTAC AAGTGGCATG 1680

ATATGGTCAA AGGCGATATC GAATTTGAAG GTGGCAATAT CGGTGGTGAC TGGGTTATCC 1740

AAAAGAAAGA CGGTTACCCA ACTTACAACT TTGCCGTTGT TATCGATGAC CACGATATGC 1800

AAATCTCTCA TGTTATCCGT GGAGATGACC ATATTGCTAA TACACCAAAA CAGCTTATGG 1860

TCTATGAAGC TCTTGGTTGG GAAGCTCCAG AGTTCGGTCA CATGACCTTG ATTATCCACT 1920

CTGAAACTG 1929

(2) INFORMATION FOR SEQ ID NO: 304:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 708 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 304:

AAATTTAAGA AAAAGGAGAC ACATCATGTC TAAAAAAGTA TTATTTATCG TCGGATCACT 60

ACGTCAAGGT TCTTTCAACC ACCAAATGGC GCTCGAAGCT GAGAAAGCAC TTGCTGGTAA 120

AGCGGAAGTT AGCTACCTTG ATTATTCAGC CCTTCCTCTC TTCAGCCAAG ATTTGGAAGT 180

TCCAACACAT CCAGCTGTAG CTGCTGCTCG TGAAGCAGTT CTCGTTGCGG ATGCTATCTG 240

GATTTTCTCT CCAGTCTACA ACTTCTCTAT CCCTGGTACA GTGAAAAACT TGCTTGACTG 300

GCTATCTCGT GCCCTTGACT TGTCTGATAC ACGTGGCGTT TCTGCCCTTC AAGACAAGTT 360

TGTCACAGTA TCATCTGTAG CCAATGCAGG GCACGATCAA CTTTTCGCTA TCTACAAAGA 420

CCTCTTGCCA TTTATCCGTA CACAAGGCGT TGGTGATTTC ACTGCTGCAC GTGTTAATGA 480

CTCTGCCTGG GCAsACGGAA AATTGGTTCT TGAAGAAACA GTCCTAAACT CACTTGAAAA 540

ACAAGCTCAA GACTTGGTCG AAGCTATCAA GTAACTAACA CTCAATAAAA ATCAAAAAGC 600

AAACTAkGAA GCTArCCGCA AGCTACTCaA gCACTGCTTT GAGGTTGTAG ATAGAACTGA 660

CGAGTGTnnA ACATATATAC GGTAAGGCGA CACTGACGTG GCTTGAAn 708

(2) INFORMATION FOR SEQ ID NO: 305:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 781 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

CTTCTTTTCT TGGAAATAGG TGTATAATAC GTTTATTAAA TTTTTGAGGA GTTGTCTATG 60

AAGAAAAGTT TTATCCATCA ACAAGAAGAA ATTTCCTTTG TCAAAAACAC TTTTACCCAG 120

TATTTGAAAG ATAAGCTAGA AGTTGTCGAA GTTCAAGGTC CTATCTTGAG TAAGGTCGGT 180

GACGGAATGC AGGACAACCT GTCTGGTGTG GAAAATCCAG TATCGGTCAA GGTTCTCCAA 240

ATCCCTGATG CTACTTATGA AGTGGTGCAC TCACTTGCTA AATGGAAACG CCACACCTTG 300

GCTCGTTTTG GCTTTGGTGA AGGAGAGGGT CTCTTTGTCC ACATGAAAGC CCTTCGTCCA 360

GATGAGGATT CCTTGGATGC AACCCACTCT GTTTATGTTG ACCAGTGGGA CTGGGAGAAG 420

GTTATCCCAA ATGGTAAGCG TAACATCGTT TATCTAAAAG AAACAGTTGA GAAGATTTAT 480

AAGGCTATTC GCCTGACTGA GCTAGCTGTT GAAGCCCGCT ATGACATCGA GTCTATCTTG 540

CCAAAACAAA TTACCTTTAT CCATACAGAA GAATTGGTAG AACGCTACCC AGACTTGACA 600

CCGAAAGAAC GTGAAAATGC GATTTGTAAA GAATTTGGAG CCGTCTTTTT GATTGGTATC 660

GGTGGCGAGT TGCCAGATGG TAAACCGCAC GATGGACGTG CACCAGACTA TGATGACTGG 720

ACAAGCGAGT CTGAGAATGG CTACAAGGGT CTAAATGGTG ATATTCTTGT CTGGAATGAG 780

T 781

(2) INFORMATION FOR SEQ ID NO: 306:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 846 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 306:

CCCGCATCTT GTAGGGTTTT AACGGGCACG ATTTTCATAT CCGTCTTGAT TGTTTTAGCC 60

GCTTCTAGGG CTGTTTGGTA GTTGTTTTTC GCGTCCGGAT GCGCCTTTTG TTCTTCTTCG 120

CTAACAGGGT TATCAGGAGC AAAGAAAATA GCAGCACCTG CCCTAGCCGA AGCTACAACC 180

TTCTTATCAA TACCTCCAAT GTCTCCCACA TTACCATCGC GGTCAATGGT ACCTGTACCG 240

GCAACAATAC GACCATTACG AAGATCTGGG TGAGCTATTT GAGTATAGAT AGCTAGACTA 300

AACATGAGAC CAGCACTTGG ACCGCCAATA CCAGCTGTTG AAAAGCTAAT TGGGACATTG 360

CTGATTACCT CTGTACGGTC AATCAAGCCG ATTCCAATTC CATTTTTGCC ATTTTCCAAG 420

GTGATGATTT TTCCTTCTGC AGACTTGGTT TGCCCATCCT CTTCATAGGT GACCTTGACG 480

GAATCCCCTA ATTTTTGAGA ACTGACGTAA TCAATCAAGT CTTTGGAACT ATCAAAGGTC 540

TGATCATTGA CTGCTGTGAC TGTATCAGAG ATATTGAGAA TCCCTTTAAA GGTTGAATTA 600

TCCGTCACAT TCAAAACATA AACTCCAAAG TACTTGAGTT CGATATCCTT ACCAGCTGTT 660

TTTAGTCCTT GATACTTGGC CATATTTTGC GATGTTTGCA TGTAGAATTG ATTGATTCGC 720

ATAAATTCAA CATCGGAAGA ACCACCTGTA GTCTCCTGAG CACTACGAAT ATCTGTAAAA 780

GGTGTCAACC AAGCATAAAT CATATGAGCT AAAGTGGCAT GTTGAACACC AACCGTAACG 840

AATTGT 846

(2) INFORMATION FOR SEQ ID NO: 307:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 829 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 307:

GCGATCTGCT TGGGCTTTTC CTATTACCTT ATCTAATAAA TAGGTACGCA GACTCATAAC 60

CATATAAAGT CCACCCCCCA TGGCACCGAC AAGAGCTACA TAAAAGAAGC TCCACAAACG 120

TCCACTTGGT TGGAAGAAAA ATCCTAACAG CCACTGGATG GTTCCTATTA ACAGAAACAT 180

GACTAGGGTC AGCAAACTGA TTAAAATGGT TCGCTTCAAA ATCACCTTGC GCTTGACACC 240

AGTTACTTTA CAAATATCCC GATACATCAA GACGTTAGGA ATGATGAGAG CAATGGTTGT 300

TGAAATCAAA GGACCATAAC TGTGGAAGAG GGCGATGGTA GGTAGTTGCA AGACTAGCTT 360

GGCAATAGAA CCATAGATAA AATAGAGAAC GGCCTTGCGG TTGCGGAACA TGGCCTGAAG 420

CATTGGAGAC AAGACCATGT ACAAGCCTAA AATAATAGAC TGCAAAACTG CAAAGACAAA 480

TAAGCCCAGA GCCAAACTAT CTGGCTTACC ATAGAAGACC GTATAAAGAG GTTCTCCTAC 540

CATAACCACT CCAACCGTTG CTGGTAGCAA GAACATAAAG AGTAGGGTGA GACTGTCCTG 600

AACGAGACGA GAAGCTGCTT TCAAGTCCCC CTTGACATAG TTTTCCGTCA AAAGTGGCAA 660

ACCAACACTC CCAATCGAAA CCCCTACAGA AATCAAAATC ATCGTGATTT TATTAGGATT 720

GGCTGAGAAA TAAGAAAACA TGACAACCAA GTCCTCATTG CTGTAGTTGG TAAACCAGCT 780

CATACTATTG ATAAAGGTCA GCTGAGTCCA AATCTGGAAG AGCTGGATG 829

(2) INFORMATION FOR SEQ ID NO: 308:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 464 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 308:

CGAACATCTT GCTGGCTGAT TCGTCTGCCG CCATCGCAGC CCCGAACACA TTGCGACCCA 60

TGGCAAGCGG GCTCAATCCG CACATGGGAT CCGTGCCAAA GCCCCGCGTG TGCATCATTT 120

GCTCATCTAG TAACGTATGA GGTTTGCCTT CGCTGTCGAT AAACCGATAT TCAATCGCAC 180

CACTGCTCGT TCTCCGCGGA GGGGAAACCG ACTGCGGTAG GATGAACTCC AGAGAAGAGA 240

GATCACGACC TACCAGGTGC GGCTCGTTGA AGCTGTTGCC GCTTAGCAGC AGGCTCGCCA 300

CCACGCATTC CCAGAACTCA ACGGGGGTTT GATCGGCGTT CGGTTGCTGA CTAATAACTC 360

GGTGCACGGG ATGCGAAGTG GCCACTTCTG GCACACCGTT CTTGTCTTCG TAGAGAGCAA 420

TTGGGAGGGT GGCCAGCGTT TCGGCGATGA GGCGCACGCA GGCC 464

(2) INFORMATION FOR SEQ ID NO: 309:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 982 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 309:

CCGTCTATAA TGGTAATAGA TTTTATTTGG AGGTTTTTAT GTCATTTCTA TCAAAAAATG 60

GAGCAGGTAT CTTGGCCTGC CTTCTCATTT CCATCCTATC TTGGTACTTA GGAGGATTCT 120

TCCCTGTGGT TGGCGCGCCC GTTTTTGCCA TTTTCATAGG CATGCTCCTA CATCCCTTTC 180

TCTCGTCCTA TAAACAACTG GATGCTGGTT TGACCTTTAG TTCCAAGAAG TTGCTCCAAT 240

ATGCCGTTGT CTTGCTTGGT TTTGGTCTCA ATATCTCGCA GGTCTTCGCA GTTGGCCAAT 300

CTTCACTCCC TGTCATCCTG TCCACTATCT CAATAGCTCT GATTATTGCC TACCTCTTCC 360

AGCGTTTCTT TGCCCTGGAT ACAAAACTGG CTACCTTGGT TGGAGTAGGT TCTTCTATCT 420

GTGGGGGTTC TGCCATTGCA GCGACAGgCC CGTTATTGAT GCTAAGGAAA AGGAAGTAGC 480

CCAAGCCATT TCCGTTATCT TTTTCTTCAA TGTCTTGGCT GCGCTCATCT TTCCAACCCT 540

CGGCACCTGG CTTCATCTAT CCAATGAAGG CTTCGCCCTC TTTGCAGGGA CTGCGGTCAA 600

CGACACTTCC TCTGTAACGG CTGCCGCCAG CGCTTGGGAC AGTCTTTACC AAAGCAATAC 660

CCTCGAGTCT GCAACCATTG TTAAACTCAC ACGTACTTTG GCCATTATCC CTATCACGCT 720

CTTTCTATCC TACTGGCAAA GTCGCCAACA AGAAAACAAG CAAAGCCTGC AACTGAAAAA 780

AGTCTTCCCA CTTTTTATCC TTTACTTTAT CCTTGCCTCT CTCCTCACTA CACTACTCAC 840

CTCTCTAGGT GTGTCCAGTA GTTTCTTTAC TCCTCTCAAA GAACTCTCTA AATTCCTTAT 900

TGTCATGGAC ATGAGTGCTA TCGGTCTCAA AACCAATCTG GTCGCTATGG TCAAATCCAG 960

TGGAAAATCC ATTCATCATG GA 982

(2) INFORMATION FOR SEQ ID NO: 310:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1939 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 310:

CTAGCTGCCA ATATGATTGG GGTGCAGAAG CGCGTGATTA TCTTTAATCT TGGCTTGGTT 60

CCTGTGGTCA TGTTTAACCC AGTGCTTCTG TCCTTTGAAG GATCCTATGA GGCAGAAGAA 120

GGCTGTTTGT CCTTGGTAGG TGTGAGATCA ACTAAGCGTT ATGAAACCAT AAGGCTTGCC 180

TATCGTGACA GCAAGTGGCA GGAACAGACC ATTACCTTGA CAGGCTTCCC AGCTCAGATT 240

TGCCAGCATG AGCTGGATCA CTTGGAAGGA CGAATCATTT AGGAGGAAAG CAAATGAAAC 300

GAATAGTCTT TGAACTTATT TTTATCGCAA CGACCTGGTA TATCTTTTTA CCGCCCCTTA 360

ACCTGACCAG CTGGGAATTT CTCTTCTTCC TCTGTGGGCA TTTGTTAGTT GTGGCAATAT 420

TATTTGGCTT TGGCAAGGGG ATAAACCTTG TCAAAACGGT TCATGTGCGC CACGGTAAGG 480

CGGAAGCTGC CTTAAATCTT GAGGGTTTCA AAATCAATCG GTTAGGGAAA ATTCTGTTAG 540

CTTCGATTGG AGGAATTCTT CTCTTGGCAG CTTTGGTTTc CTTGGTAACT TCCAGCATGT 600

TTCAGGCTAA AAATTATGCC AATGTAGTCA CGGTTACGGA AAAAGACTTT ACTGAATTTC 660

CTAAGAGTGA CACCAGTAAG GTTCCTATCC TAGATAGAAG TACTGCTGAA AAAATTGGAG 720

ACCGCTACTT GGGTTCCCTA ACCGATAAGG TGTCGCAATA CGTAGCGGCA GATACCTATA 780

CCCAATTGAC AATTGATGGG AAACCTTATC GGGTCACACC ACTAGAATAT GCAGACCCTA 840

TCAAATGGTT TAACAATCAA GCCAAGGGAA TCGGTGAGTA TATTAAGGTG GACATGGTAA 900

CTGGAAATGC GGATTTGGTG GACTTGAAGA CACCAATCAA GTATTCAGAC TCGGAGTATT 960

TTAACCGTGA TGTCAAACGT CACCTGCGCT TGAAGTACCC GACCAAAATC TTTAAAACTC 1020

CATCTTTTGA GGTGGACGAT GAGGGCAATC CTTTCTATGT AGCAACGGTT TACCAAAAGC 1080

AATTTGGACT TGCTGTTCCT CGTCCTGCTT CAGTCATTAT CTTGGATGCT ACAAATGGAG 1140

AAACCAAGGA ATACAGCTTA TCAGATGTTC CAGAATGGGT GGACAGGATC TATCCAGCAG 1200

AGGAAACCAT TGAGCAAATC AACTACAACG GCAAGTACAA GGACGGTTTC TTGAATGCCA 1260

TGATTTCCAA GAAAAACGTG ACCCAGACTA CCAATGGCTA TAATTACTTG TCTATCGGTA 1320

ATGACATCTA TCTCTACACA GGTGTGACGT CGGCTAATGC GGATGAGAGT AATCTTGGTT 1380

TCATCCTTGA AAATATGCGA ACAGGAGAAA TCACTAAGTA TAGCTTGGCT TCTGCGACAG 1440

AAGAATCAGC CCGTGAATCA GCAGAAGGTG CTGTTCAGGA GAAATCCTAC AAAGCAACCT 1500

TCCCAATCCT CATCAACCTC AATGACAAGC CTCTCTACAT CATGGGCTTG AAGGACAATG 1560

CTGGCTTGGT CAAAGAGTAC GCCCTGGTAG ACGCAGTCGA GTACCAAAAT GTTATCGTTG 1620

CTACTACAGT GGAAGAGATG CTCAGCAAGT ATGCCAATAA AAACGACCTT GAAATTGACA 1680

ATGCAACGAC AGAAAGCATC AATGGAGTAG TAGCAGACCT CAAATCAGCT GTTATCAAGG 1740

GAGACACTGT CTACTTCTTT AAAGTTGATG GCAACATCTA CAAGGTCAAG GCTTCAGTAT 1800

CCGATGACCT TCCTTACCTT GAAAATGGTA AAACCTTCGA AGGTCAAGTA GGAAAAGACA 1860

ATTATCTCAA GACCTTTAAG CTACGGTAAA AATAGGTTTT TTTCAGAAAG TATATGTTAT 1920

AATAAGGTAA ATTAAGCCG 1939

(2) INFORMATION FOR SEQ ID NO: 311:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 907 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

CCTGCTAATA GAGAGAAAGA CTAGGAGTAG AAGTAAGCCA ATTAAATAAT GAGAAAGTTT 60

CATACCCCGT CCTTTCATGT AGATTTGGTA TCGAAAGATA TCTGCGGATA TAAATGTAAC 120

ATTATTTTTC TAATCTGTCA ATAAAATTTC TGACAATTTA ATAAATACAA CAAGGAGAGA 180

GCAACAAGAC TTTCTCCTTT GTTATCCTAT TCTAAAATGT TTTTACCTTA ATCTGATAAA 240

ATAATATCTT CGAGGGAGTA GCTAGCCGTC CAATCAAGAT ATTGTTTAGC TTTTGAAGCA 300

TCTGCTAGGA CACTGGCTGG GTCACTAGCA CGTCGAGCAA CAATCTCGTG TGGGATTTTT 360

TAATTTAGTA ATTCTTCAGC AGTTTTAAAG ATTTCTTTGA TAGTATAGCC TTTTTTAGTT 420

CCTAAGTTAA AGATTTGAGA AGAACTGTCT TCTTGAAATA GGTAGTTCAT TCCTTTAACA 480

TGAGCCTATG CAAGGTCCAA GACATAAATG TAATCTCGAA TACATGAACC GTCACGTGTA 540

TCGTAGTCAT CTCCAAATAT TTTTAAGCTA TCATTTTGTC CCAATGCGGT CTTGTTGATA 600

TTTGGAATGA TGTGAGTTGG ATTTTTCACA CGCAGACCGT TTGAAGCATC CATTTCAGCC 660

CCAGCAACAT TAAAGTAACG GAAAATAACA TATTTCCAGT CGTAGCGATT GGCCATCCAG 720

TAAATCATTC GTTCGCCCAT CAGTTTTGTC TCTGCATAAG GGTTGACAGG GTCGAGCAGG 780

GTATCTTCAG TCACCGGCTT GTCAATACAG TTATTTCCAT AGAGAGAAGC AGTCGAAGAG 840

AACATGATTT TTTGAATGCC AACTTCAGAT AAGACTTTGA GAACTTGGTT CATACCAGCA 900

ACGTTGG 907

(2) INFORMATION FOR SEQ ID NO: 312:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2170 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 312:

CCACATAAAG GTAAATATCT TTTGTACTAT CTTGGGCATC CAAGAAAAGC AATTGGGCAA 60

TAACAGAGTT AGCCATATTG TCTTCAACCG GACCTGTCAG CATAATGATG CGGTCTTTGA 120

GAAGACGTGA GTAAATATCG TAAGAACGTT CTCCACGGCT TGTTTGTTCA ATAACTACAG 180

GAATCATTCA TTTCTCCTTT TGAGTTTTAA TTTTGTTGGT CAAATGACTG AAGATAAGAC 240

1340

TATTATAATA TCTTGGTCAA AAAAGGTCAA ATTTTTGCTC TGCTTTCATT AGACAGAAAC 300

AAAAACCCAA CCTCCTTTCG TGACTGGAAA TACTTTTCCA AGTCATTCTT CTTTTCGATC 360

TTATTTTGTA CCGAACAAGC GGTCTCCAGC ATCTCCAAGA CCTGGAACGA TATAACCGTG 420

TTCGTTCAAA CGTTCATCCA AGGCTGCTGT AAAGATTTCT ACATCTGGAT GAGCTTCTTG 480

AAGGGCTTTT ACACCCTCTG GAGCAGATAC AAGGCAGACA AATTTGATAT TTGATGCGCC 540

ACGTTTTTTA AGAGAATCAA CAGCCAAGAT TGCTGAGCCA CCTGTTGCCA ACATTGGGTC 600

TACTACAAAA ATTTGACGTT GGTCAATGTC CTCAGGCAAT TTCACCAAGT ATTCAACTGG 660

TTGAAGTGTT TCTTCATCAC GGTACATACC GATGTGGCCA ACTTTAGCAG CTGGAACCAA 720

GTTCAAGAGA CCATCAACCA TCCCGATACC TGCACGCAAG ATTGGGACGA TGGCCAATTT 780

CTTACCTGCC AATTGTTTTT GAACTGTTTT TGTAATTGGT GTTTCGATTT CCACATCTTC 840

TAGTGGAAGA TCACGAAGTA CTTCATACCC CATCAACATT GCAATCTCAT CTACTAGCTC 900

ACGAAAAGCT TTTGTAGAAG TATCTGTACG ACGCAAGATT GACAATTTGT GTTGAATCAG 960

TGGGTGATTA ATAACTTCAA TTTTTCCCAT TTTTGGAATT CCTTCTTTCA ATTTATTCTT 1020

CTTATTATAC CAAAAAACGG TTTAAAAATC TTTCTAAACC ATTTATTTTT GATAATTTTT 1080

ACATTAGATC AGCCTCTTTA AGAGCTGTCT GTACTGTCTC AAGTGGTAAA TGGGTCAATT 1140

CTGTCCCTTT TTCTTGATAA AGGTATTGGG CGTAGTCGTC CATTCGGTAC TGGTTGATAT 1200

AAACCACGCG CTTGCAGCCG ACCTGAAGCA ATTGTTTTGT ACAGTTGAGA CAAGGAAAAT 1260

GGGTTACATA GGCTGTAAAG CCTTTGGGAA CACCACGCTC AGCACCTTGA AGGATAGCAT 1320

TGACCTCAGC GTGAAGGGTG CGAACGCAGT GGCCTTCAAT GACCAAACAT TCGTGATCAA 1380

TACAATGCTC AGTCCCTGAC ACCGAACCAT TGTAACCAGT GGAAATAACC TTATTATCTT 1440

TTACCAGAAT CGCGCCCACT TTAGCACGTT TACAAGTGGA ACGATTCGCA ATTAGTAGAG 1500

CTTGGGCTGC AAAATACTCA TCCCAGGCCA GTCTTTTTTC AGTCATCTCT TTTCTCCTTT 1560

TTCTCTATTT TTTAAAAAAT GGTAAACCTA AATCTGCAAT CTTTTCAGCT GGTACCTTCA 1620

TGCCATCCTT GATCCATTTT AGAAGGACAG AGACGATGGC TGAGCTCCAG AAGGAATGAA 1680

GATAAGAGCT GACACCTTTT GATTTCCCAT GGTATTTTTC TAGAAATTCC TGCATGGCTT 1740

GGACAAAGAT TTTTTCCAGA TGGTAATCCA AGGCCAATTG AATTACTCTA GCTTCCTTTC 1800

TGGCCTCCCG GAAAAGGTGA ACCCAAACCA AATAAAGGTC TGTCTTTAAA TCGTAATGAT 1860

GCAGCTGTTC CATAATATTG TGGACAGTTC GTTTAAAGAC GCTCTCTAAA ATTTCCTCTT 1920

TGGAGTCATA ATTGCGATAA AAGGCCGCAC GCGAAACACC TGCACGTTTG ACCAATTCAG 1980

AAATACTAAT CTTGGTCAGT TCCTTTTTTT CCAAGAGTTG CAAGAGGGCT GTTTCAATGG 2040

CTTCTCTGGT TAATAAATTG GATTCTTGGT TTGATTTTCT GAGATTTTCA AGAGACTTTT 2100

CAGAGATTCT ACGTTCAGAC ATAACATTTT CTTTCTACTT GTCACAACAG ACGGATGATG 2160

CTTTTGTTTC 2170

(2) INFORMATION FOR SEQ ID NO: 313:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 539 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 313:

ATCTGCACGA ATCAGGGCTT TCTAAGTGAC TATTTCCACC GAAATATTAT TTATATCAGG 60

AGGACATTCA TATGTCACGT TATACAGGAC CATCTTGGAA ACAAGCTCGT CGTCTTGGCC 120

TTTCACTTAC AGGTACAGGT AAAGAATTGG CACGTCGTAA CTACGTACCA GGACAACACG 180

GACCAAACAA CCGTTCTAAA TTGTCAGAAT ACGGTTTGCA ATTGGCTGAA AAACAAAAAC 240

TTCGTTTCAC TTACGGTGTA GGTGAAAAAC AATTCCGTAA CTTGTTCGTA CAAGCTACAA 300

AAATCAAAGG CGGAATCCTA GGTTTCAACT TTATGCTTCT TTTGGAACGT CGTTTGGATA 360

ACGTTGTTTA CCGTCTTGGT CTCGCGACTA CTCGTCGTCA AGCTCGTCAA TTCGTAAACC 420

ACGGTCACAT CCTTGTTGAC GGGAAACGCG TTGATATCCC ATCATrCCGC GTAACTCCAG 480

GTCAAGTGAT CTCAGTTCGT GAAArATCAT TGAAAGTTCC AGCAATCCTT GAAGCAGTA 539

(2) INFORMATION FOR SEQ ID NO: 314:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 667 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

CCGGTTTTGC TCCTTCTCTA CGGCTACGAC GTGATGTATC TCTGATGATA TCCACTGTTT 60

CTGTAGCAGG CGTAGGTGTT TCTGGACCTG CTTGTTCTGC TTTTTTCTCT GCCGTCGTAT 120

AGGAAACAGC TACCCTTGTT GGGGTTTCAT TGTATTCTCT TTCAAGTTTC TTAGGTCTAA 180

CAGGACCTGG ACCTGGTCTT GATCCACTTT CTTCCGCTGG AGAAGAAGGT ACATCTTGAC 240

TTGGATGACT TGGAACACCA GGAGTTTCTC TTTGAATCTC ATCTGCTGGA GAAGCTGGTA 300

CACCTTGACT TGGGTGAGTA GGCACGGTAG GAGCTTTTCT CATAATCTCC TCTACCGTTG 360

ACAAGGAATC AGCCATGAGT TCTTCAGTTG AAGGTTCATT TGCAGGAGTG CGAACTACTG 420

CCTCATCTTC TTTCAGAACT TCATCATAGC CTTTTACTTT TTCTAAATCT CTCAGAATCT 480

GCTCTTTAAA GCGTAATTTC TCTTCTGCTC TTGACTTTTC ACTCAAAAGT TTTTCCTCCT 540

TGTTGAGAAT CCATAATATT AGAGCTGAGA AGTCCAAAAA AAGCAATCTA TGATACTTTT 600

CCTAACGGAT TTTGTCATTT CCCAGACCAT ATCATACCAT GTTTCCCCTG CAAAGGTTGA 660

CTGGGAA 667

(2) INFORMATION FOR SEQ ID NO: 315:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1483 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 315:

GGGAAGCCAA GGTATTTTAT CGGATGAAGT TGTTACTAGT TCTTCACCGA TGGCTACAAA 60

AGAGTCTTCT AATGCAATTA CTAATGATTT AGATAATTCA CCAACTGTTA ATCAGAATCG 120

TTCTGCTGAA ATGATTGCCT CTAATTCAAC CACTAATGGT TTAGATAATT CGTTAAGTGT 180

TAATAGTATC AGCTCTAATG GTACTATTCG TTCCAATTCA CAATTAGACA ACAGAACAGT 240

TGAATCTACA GTAACATCTA CTAATGAAAA TAAGAGTTAT AAGGAAGATG TTATAAGTGA 300

CAGAATTATC AAAAAAGAAT TTGAAGATAC TGCTTTAAGT GTAAAAGATT ATGGTGCGGT 360

AGGTGATGGG ATTCATGATG ATCGACAAGC AATTCAAGAT GCAATAGATG CTGCAGCTCA 420

AGGGCTAGGT GGAGGAAATG TATATTTTCC TGAAGGAACT TATTTAGTAA AAGAAATTGT 480

TTTTTTAAAA AGTCATACAC ACTTAGAATT GAATGAGAAA GCTACAATTC TAAATGGTAT 540

AAATATTAAG AATCACCCTT CCATTGTTTT TATGACAGGT TTATTTACGG ATGATGGTGC 600

GCAAGTAGAA TGGGGCCCAA CAGAAGATAT TAGTTATTCT GGTGGTACGA TTGATATGAA 660

CGGTGCTTTG AATGAAGAAG GAACTAAAGC AAAAAATCTA CCACTTATAA ATTCTTCAGG 720

TGCATTTGCT ATTGGGAATT CAAATAACGT AACTATAAAA AATGTAACAT TCAAGGATAG 780

TTATCAAGGG CATGCTATTC AAATTGCAGG TTCGAAAAAT GTATTAGTTG ATAATTCTCG 840

TTTTCTTGGG CAAGCCTTAC CCAAAACGAT GAAGGATGGG CAAATCATAA GTAAGGAGAG 900

CATTCAGATT GAACCATTAA CTAGAAAAGG TTTTCCTTAT GCCTTGAATG ATGATGGGAA 960

AAAATCTGAA AATGTGACTA TTCAAAATTC CTATTTTGGC AAAAGTGATA AATCTGGGGA 1020

ATTAGTAACA GCAATTGGCA CACACTATCA AACATTGTCG ACACAGAACC CCTCTAATAT 1080

TAAAATTCAA AATAATCATT TTGATAACAT GATGTATGCA GGTGTACGTT TTACAGGATT 1140

CACTGATGTA TTAATCAAAG GAAATCGCTT TGATAAGAAA GTTAAAGGAG AGAGTGTACA 1200

TTATCGAGAA AGCGGAGCAG CTTTAGTAAA TGCTTATAGC TATAAAAACA CTAAAGACCT 1260

ATTAGATTTA AATAAACAGG TGGTTATCGC CGAAAATATA TTTAATATTG CCGATCCTAA 1320

AACAAAAGCG ATACGAGTTG CAAAAGATAG TGCAGAaTwT TTAGGAAAAG TATCAGATAT 1380

TACTGTAACA AAAAATGTAA TTAATAATAA TTCTAAGGAA ACAGAACAAC CAAATATTGA 1440

ATTATTACGA GTTAGTGATA ATTTAGTAGT CTCAGAGAAT AGT 1483

(2) INFORMATION FOR SEQ ID NO: 316:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2453 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 316:

CCTGAACGCT TTTTTATAAA TATCATAAAG CCAATCTGAT TTATCAAGTG TGTCTAAGCG 60

ACGCGAATTA AAATTCATTG CATACTCCAT CGCTTCTAAA AAACTCATTT TTGAAAAGAC 120

GTTAAAATCA TCTAAATTCT GACTCCAATA TAATAACAAA ACCAATCCCA TAATATCCTC 180

TGGTTGATTA TTCAATAAAT TTAAGTTGGT TTCATAAAAC CCTGGAGTTC CAAATAGAGG 240

CAACTTTTTT TCTTCAATTT GAGTTTCTTT CCTTAGGGCA TGCTCAAAGT CTATAATATA 300

AATATTATTT CTATTATCAA TAAGTATATT ATTAAATGAT AAATCTCTAT AGGAAAGATT 360

ATATTTGGAG TTTATTATCT CCATATAATC AATTAATGTT AAAAACCAAT CATACGAGCC 420

ACTAACCATA TTATACTCGC TTAATTTATC TGCAATAATA AACTCAAATT CCACAAAATA 480

CGAATTCTTT ATGTAAAAAT CGTTAAAAAC TTTTGGAGTA AATTCCTCCT TTTCCAATTC 540

TACTAATATT TCTCTTTCAT TTATTAAACG ATTCACAGAA TCTCTATTTG TAAAATCAAC 600

CAACGATAAA TCACTAGCTT CTTTTAATAA AGAATAAACT CGCTTTTGAG TATTAAATAC 660

TTTATAAACT CCACCTTTGG CATTTTTAGA AATCACTTCC AAAATAATAT ATTGATCAGG 720

AATAGTGTTA TATCTTGGAA TATAGTAATC CCTTATTGGA ACATTCACAT TTGAAGGGAT 780

TTTCTTATCT CTTTTATCCT TGAAAGTGCT ATCTTTTACG AACTCCCCAT ATCTGTAATA 840

TACAACCTCG CTAAGTTGAA ATCTGAAATC TGATGGTATG TTTACACCCT TTACACCTTT 900

ATACAATATT TCTAATTTGT GTAACAAACG TTGAAACTCT TTATTATCTT TTGGATAAAT 960

TGTAATGAAT TTCCCGACTT GTGAATAACC ATTAAGCCCT GTATTTTGCA AAGAAAGTTC 1020

TTTAATGCTA ACCAAAAATT TGAAATTTAT CTTCTTCTCT CTAGAAAATA TAAAATCAAA 1080

GAATTTTTTA GCAACCAAAT TAGCATTTAA TATTGAAGCG CTCAGGTGTA TTTTAAATCC 1140

CTTAGATTGG GTGATATTAG ACGGCAAATT ATATAACCAA TGTTCATCAC TAAAATTATC 1200

ACTAATTTTA TATTCTAATA ATAAATTATG GTATGCGTCT TCTATTTCAG TTTCATAGTC 1260

CAAATAGTTT AAATACTTTT CGTAATTCAT ATTAAGAAAT CTTCTCCATA AATTTTTAGA 1320

CCATCATTTA AAGCCAAACA ATTTAAAGCG TGATAATAAA TGTTGATAAT CAATGTAACT 1380

TTCAGTCCTC TATTTTGTAA TTCCTTCACC AATAATTTTA TGCTATATCT ATTTTCTCGA 1440

GGCAATTTAT AGGACTTCAA GATAAAACCA TAAAAGAGAT AAGTATTATA ATCTGACAAT 1500

CCAGTTTCAG AATAATTTTT TAGAAAAATA TCTAGTGATT CTGATAATTC ATCCGGAATA 1560

ATTCTTTTAA CATCGTATTT ATTTTTCATA TCGGCCACTC TTCCTTAAAA AGCTCACAAT 1620

AAAATTTTAA ATTTCTATAC AACAATCCGA GAGTAGTCTC ACAATTTGAA CATTTCACAT 1680

CACTCTTAAT ATATAAAAAA TGAATTAATC AGAAACCTCT GACTAAGATT TCCTAATTAA 1740

TTCACTTTCT ATATCATAGT AAGGAATTCT ATTATCCCTA ATTGAAAATT GAAATTTTAT 1800

GTTTTATATA TTAACAATTA TGCGGATTGT AAATCTTGTC TAACAAAATG GCAAGTGCTA 1860

CTATGTGCCC CAGAAGGCGA TGCAACGCTA TTTTGAATTG AAAGAGCATA ATCATCCATA 1920

TCATTTAAGT CACGGATTAG CAATGCTTCC TTCTCTCTTC CGACAATTCC AAATTTTCTA 1980

ATTACCTTTT CAGGATTATC AAAAAATTCT CCAACAACTT CCATATTTCC TTGAAGTTCA 2040

TTCAAGAAAG CTTTCATTTG ACTACTCATT ATATAGCTCC TTTTCTATTA CTTTATTTGG 2100

AATCAAAACT TACTTGTACA TTGGAAACAC CTCTATTCTA CGCTTTCATA TTGCTGCATG 2160

ACACTTTCAA AATCAAATTG CTAAAAATAA TTTTTTAAAG CTTAATTTAG ATTTAATTAC 2220

ATATATCTCA AAAAATTGTT TTGAAATTAG TAAATTAAAA TAGGTTTCTG TACTTATAGG 2280

AACTAGTTAT AAAAACTTCG CCCATCATAA AATATCTATT TAAGTAAAAC AAAAATTTTA 2340

TAATTTTTTG ATTTTTAAGT GACTATAATC TCCTATCTAT AAATACCATT CGCAGGACCT 2400

GGATCAATCC CTCTAGCCAT CTTATGAACT TGAGTTCCTC CAGACAGTCC CGG 2453

(2) INFORMATION FOR SEQ ID NO: 317:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1049 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 317:

CCAATTTGAA GGCTCTAAAA CAATGGAAAA GTGCTACACA GATGTGACAG AATTTGCCAT 60

TCCAGCAgTA CTCAAAAACT TTACTTATCA CCAGTTTTAG ATGGCTTTAA CAGCGAAATT 120

ATTGCTTTTA ATCTTTCTTG TTCGCCTAAT TTAGAATAAG TACAAACAAT GTTGGAACAG 180

GCATTCAAAG AGAAGCACTA TGAGAATACG ATTCTCCATA GTGACCAAGG CTGGCAATAC 240

CAACACGATT CTTATCATCG GTTCCTAGAG AGTAAGGGAA TTCAAGCATC CATGTCACGC 300

AAGGGCAACA GCCCAGACAA CGGCATGATG GAATCTTTCT TTGGCATTTT GAAATCGGAG 360

ATGTTTTATG GTTATGAGAA GAACTTTAGA TCTTTAGAAA ACCTTGAACA AGCTATTGTG 420

GACTACATTG ATTATTACAA CAACAAGAGA ATTAAGGTAA AGCTAAAAGG ACTTAGCCCT 480

GTGCAATACA GAACTAAATC CTTCGGATAA ATTAATTGTC TAACTTTTGG GGTGCAGTAC 540

ATTTTTGGTA TATATAAAAT TTGTAGGAGC TATATCTACA ATTTTATATT CCCAGTTTAT 600

GGATGTAACT TACTATATTC ACAATGTTAT CCAGTGTTTT TTCTCTAATA TTTAAGGAGT 660

GTTCTGTTTC TCGAATAAAT TCTTCAAAGT TTAACCCGTC AACTTGTTCC TGAACAAGAA 720

AATAATCATC CACGATATAA AATTCATCAG TTAAATTAGT AGTATAACTT TTATCGGCTA 780

ATTTTTTTAG CATGTGAGCT TCATTTTTTA TATCATCAAG AGCTGTCCAT TCTCCTTCAG 840

CATCATAATT CACAAAAGGT CTTGACTGCT TGATGATTAC TTTTTGCCCG TCCGATTTTC 900

TAATTGCCCG ATAAACATTT CCTTTATTTG ATCTCTTAAT AATTTTTTCC ATTTTGTATT 960

TATTTATTGC AGAGTCCTTA CTTGAAACTT CACATGTGGT TTGAAAATAA ATCCTTTTTT 1020

CTTCTTCTGA AAATAAATCC ATTTTCCGG 1049

(2) INFORMATION FOR SEQ ID NO: 318:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 776 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 318:

TTAGTTGGTT AGAATCAGAA AATCGCCGAA GTGGTTATTT ATTTTTGAAT AAATTTAACG 60

AACCAATTAC AGCAAGAGGA GTTGCTCAAC AGTTAAAAAA TTATGCTGAT AAATACAAAA 120

TGAATCCTAA AGTAATTTAC CCTCATTCTT TTAGGCATTT ATTTGCTAAG AATTTTTTAG 180

CGAAGTATAA TGATATTGCC TTGCTTGCAG ATTTGATGGG ACACGAAAGT ATAGAAACTA 240

CTCGAATTTA TCTAAGGAAA ACAGCTACTG AACAACAAAA TATTGTAGAT AAAATTGTTA 300

ATTGGTAAAA AATAACAGGT GGTCAAACTG ACTACCTGCT ATTTTTGTGA TTATGGCTCT 360

TATTATGGGA ATATACCTAT GAATTGGGTT GTTATAAAAA TAAAAGATAT TTTTTCAATA 420

AATACAGGTC TTTCTTACAA GAAGGGCGAT TTAAGCATTA ATAATAAAGG TGTTAGAATT 480

ATACGTGGTG GTAATATTAA GCCTTTAGAA TTTTCTCTGT TGGATAATGA TTACTACATT 540

GATACACAAT TCATCTCCTC TGAGCAAGTT TATTTAAAAC ATAATCAGCT AATAACACCT 600

GTATCAACCT CTTTAGAACA TATTGGAAAG TTTGCAAGAA TCGAGAAAGA CTATGATGGT 660

GTTGTGGCTG GTGGATGTAT TTTCCAATTA ACACCATTCG AAAGTGCAGA GATGATGTCA 720

AAATGTCTAT TATGTAACTT GTCCTCTCCG TTATTTTATA AACAATTGAA AGCAAT 776

(2) INFORMATION FOR SEQ ID NO: 319:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 658 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 319:

TGCAATGCGG CGGCTGCATA CGCTTGATCC GGCTACCTGC CCATTCGACC ACCAAGCGAA 60

ACATCGCATC GAGCGAGCAC GTACTCGGAT GGAAGCCGGT CTTGTCGATC AGGATGATCT 120

GGACGAAGAG CATCAGGGGC TCGCGCCACC GAACTGTTCG CCAGGCTCAA GGCGCGCATG 180

CCCGACGGCG AGGATCTCGT CGTGACCCAT GGCGATGCCT GCTTGCCGAA TATCATGGTG 240

GAAAATGGCC GCTTTTCTGG ATTCATCGAC TGTGGCCGGC TGGGTGTGGC GGACCGCTAT 300

CAGGACATAG CGTTGGCTAC CCGTGATATT GCTGAAGAGC TTGGCGGCGA ATGGGCTGAC 360

CGCTTCCTCG TGCTTTACGG TATCGCCGCT CCCGATTCGC AGCGCATCGC CTTCTATCGC 420

CTTCTTGACG AGTTCTTCTG AGCGGGACTC TGGGGTTCGA TGTCGACAGC CCGCCTAATG 480

AGCGGGCTTT TTTTTCCTGA GGCTGGACGA CCTCGCGGAG TTCTACCGGC AGTGCAAATC 540

CGTCGGCATC CAGGAAACCA GCAGCGGCTA TCCGCGCATC CATGCCCCCG AACTGCAGGA 600

GTGGGGAGGC ACGATGGCCG CTTTGGTCCC GGATCAATTC GCGCGACCGG ATCGATCC 658

(2) INFORMATION FOR SEQ ID NO: 320:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1475 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 320:

CCGGCTTAAT TTTTAGAAAA CGTGGGCAGG GAACCTTTGT TCTCTCTCGT GGCAGCTCAA 60

AAAGAAAATT AATCGTTCCA GAAAGAGATA TCCGGGGACT GACAAAAATA TCTGAAGATG 120

CTCATTCTAC AATTGACTCG AGGATTATTC ACTTCAAATT AGAATTTGCA AATGAATTTT 180

TAGCAGAAAA ACTACAGGTC GCTTTGCAGA GTCCAGTTTA TAATATTTAC CGCCTGCGTA 240

TTATTGACGG TAAACCTTAT GTTCTGGAAC AAACTTATAT GAGTACCGAT GTTATTCCAG 300

GTATTACTGA AGATATTTTA CAAAAATCGA TTTACAATTA CATTGAAGGA AAGTTAGGAT 360

TGCATATTGC CAGTGCTACA AAAATCTTAC GAGCTTCTTC TAGTTCAGAA AATGAGCAAC 420

ATTACTTGCA GCTCCTTCCA ACGGAACCGG TATTTGAAGT AGAACAAGTG GCTTATTTGG 480

ATAACGGAAC TCCGTTTGAG TACTCGATTA GTCGTCATCG CTATGATTTA TTTGAATTTA 540

ATTCTTTTGC ATTACGACAT TCCTCCTAGG AGAAAATGTG AAAATGAAGC CAATCTTTTA 600

CAGACTCTAG TTTAAGAAAA ATTTAAAACA GGGCAAGAAG GTCCCATCTA TGCTTAAATG 660

GTTTCTCTTT TCTAAATAAG ATGGCTTTAA AAGAGTGATC GTTGTATCCA TCATGTTGAA 720

AAATATCTTC GTATAGCTTA TAGAGTAGGT ACTGAAATTG TTCACCTGAT CTACTTCTTA 780

TAGTTATTTA GTTTTAAATA GTGTTTCAAA CATTCTTACA CTGACGAGAA GTTTTTGAGT 840

CTTTTCTTGT AACACATATA GTATACTGTG GTTAGAATAG TAGACTGTGA CTTCTAACAA 900

ATTGCTAGAA ATGAATTTCA ATCTCCCAAT TTATTTGTTC ATATCTTCTT TTAATATATT 960

AAATAAATTC TAAATCATAA TCATTTAAAA AAATTTTATT TTTTATTTTT CATTACGAAT 1020

AATATAGATG AAGGGGAAAG AGTATGAAAA CAGAACTGTT TCTTTTGCTA TTAGTTCAAA 1080

AGGAGAAAAA ATGAAAGTAG AAAATATTTC GTATAGGGTG GATCATCGTA AATTGTTTGA 1140

TAATATTTCT TTTGATACTT CGAGTTCAGA CGTGACATTA ATTACTGGTA AAAATGGTAC 1200

AGGAAAGTCA ACTTTACTAT AGTAGATTGA AACTAGAATA GTACACATCT ACTTCTAAAA 1260

TATTGTTAGA AATCGATTTG ACTATCCTGA TCTATTTGTC CTGTTCTTAT TTCATTTCAC 1320

TATATCTCAA ATTGAGTATG ACGAAGTGCG CTCCCATGTC CTGGGAACGC ACTTTCTTCA 1380

TATTTTTCAT ATTCTTGAAT CCATCGATAA AGACTATTGG GATGAATTTT TAAAGTTGAA 1440

CTAATCATTT TTACAGGATG AGATTTACAG CAGAG 1475

(2) INFORMATION FOR SEQ ID NO: 321:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 560 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 321:

GAAATATATA TACTTCATCT TAATAGTGAG CAAGCTAAAC TTAGCATTTC ATGCCCTCAT 60

ATGGGATGTT CTTTGACTAA ATAATATGAT TATCGAGATA TATCTGGATA AATGAACTAA 120

TAAGTCTGAC GCGTAGACTT ATCAAAGTCA TTGGCATACA CCACTATGAA CTCGTTGGTC 180

TGTTCAAATC CCAACACATT ACCTGAGAAG AAAGTTGCAA TGTTGTTTTT GGTGCGGGTT 240

TGAATTTAAA AAATTTGTTA TGTAGTACCT AATCTAAGGA ATTAGAACAA TGCCTCTAAT 300

TTTTCTTTAA TACACTGAAA CATTGATGAT TCTGGCTGTA TTTTTGAAAC AGCTCTTCTT 360

TGCTCCTGGA AAATATCTTC AGAAGTTATA TTCTCTATTC CTAACGCTAC TTGAGTTTTT 420

TTTCTAAAAT ATTCTTTTCC GTTGCCATCT TTAGAAAAAT CATAACCTTC CCTATCTACG 480

CTGTTACACA AATTAGCTAA AAAArACTCT GGGGTTGGGA AAGGAAGATA AGAAaCGTAT 540

TTAGCCCATA ATCTATAAAG 560

(2) INFORMATION FOR SEQ ID NO: 322:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 643 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 322:

CCGCCCGGCC ACCGCTGCCT ATCCTCGGGA GAGGGTCACC TGGAGTGAAC CTAGAACGAT 60

AGACACGGTG CGGTACGACC TCGTACTACT TTCGCCGACG GCCTCGTCCG TTGTCATCCA 120

CGAACTGATC GGACATGGGT GCGAACACTT CAGAGAAAAA ATCGTTGGAC TGCGTGTCGG 180

GCCTGAGGAA CTACGGGTGG TGGCTTTTCC GAAGAACGGC TCCGGGTTTG ATGACGAGGG 240

TACACCCTCC GAAGAGATTG TACTTGTGGA GAACGGCATT GTGAGGCACG CTGTCAGGGA 300

TCGGGCGACT GGAGGAATGG CGCCTTTTTC CGGTTTGACC AAAGTGGCAT CACATGGTGT 360

CAAACCTGGC TCAAGATGTA CGCATCTCAA GGCGGAAGGG GAATCGTCAC AGGAAGGAGT 420

TACCGGAGTA CCCGCCGAAC GCACCGTTTG GATAGAGCAT TTTTCTGCAG CGAACTACCA 480

TTCAGGTCGA GCCTTTTTCA GGTCTGGCCT TGCCTGGGTA GGCAGCCGAG AAGAACTCTT 540

ATATCCCTTA ATGCCTTTCA CCATGTCAAT TGATATCTAC GAACTGGCCA GCTTATTGTG 600

GCATTTAGAC GGTCAAACGG AACGAGCACG TAGGGTACTG TGC 643

(2) INFORMATION FOR SEQ ID NO: 323:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 780 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 323:

GGTACCCACT CATTCTTGAT GAATTGTGAA CAGTTGCCCT TGGGTCGTTT TGCGAGTTGA 60

AGTCAAGAAG AGGAAAAAAA CAAAAAGGAG AAATACTCAT GGCAGTAATT TCAATGAAAC 120

AACTTCTTGA GGCTGGTGTA CACTTTGGTC ACCAAACTCG TCGCTGGAAT CCTAAGATGG 180

CTAAGTACAT CTTTACTGAA CGTAACGGAA TCCACGTTAT CGACTTGCAA CAAACTGTAA 240

AATACGCTGA CCAAGCATAC GACTTCATGC GTGATGCAGC AGCTAACGAT GCAGTTGTAT 300

TGTTCGTTGG TACTAAGAAA CAAGCAGCTG ATGCAGTTGC TGAAGAAGCA GTACGTTCAG 360

GTCAATACTT CATCAACCAC CGTTGGTTGG GTGGAACTCT TACAAACTGG GGAACAATCC 420

AAAAACGTAT CGCTCGTTTG AAAGAAATTA AACGTATGGA AGAAGATGGA ACTTTCGAAG 480

TTCTTCCTAA GAAAGAAGTT GCACTTCTTA ACAAACAACG TGCGCGTCTT GAAAAATTCT 540

TGGGCGGTAT CGAAGATATG CCTCGTATCC CAGATGTGAT GTACGTAttG ACCCACATAA 600

AGAGCAAATC GCTGTTAAAG AAGCTAAAAA ATTGGGAATC CCAGTTGTAG CGATGGTTGA 660

CACCAATACT GATCCAGATG ATATCGATGT AATCATCCCA GCTAACGATG ACGCTATCCG 720

TGCTGTTAAA TTGATCACAG CTAAATTGGC TGACGCTATT ATCGAAGGAC GTCAAGGTGT 780

(2) INFORMATION FOR SEQ ID NO: 324:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 624 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 324:

CGGGAAAAAT CAGATTGTGG GTTCAGATAT CGAATTAGCC AAGGCTATCG CAACAAAACT 60

AGGTGTCGAA TTGGAACTAT CTCCCATGAG TTTTGATAAT GTACTGGCTA GTGTTCAATC 120

1350

AGGAAAAGCC GACCTTGCCA TATCAGGTGT TTCTAAGACA GATGAACGGA GCAAGGTGTT 180

TGACTTTTCC ATTCCCTACT ATACTGCAAA AAATAAACTC ATTGTCAAAA AATCTGACTT 240

GACTACTTAT CAGTCTGTAA ACGACTTGGC GCAGAAAAAG GTTGGAGCGC AGAAAGGTTC 300

GATTCAAGAG ACGATGGCGA AAGATTTGCT ACAAAATTCT TCCCTCGTAT CTCTGCCTAA 360

AAATGGGAAT TTAATCACAG ATTTAAAATC AGGACAAGTG GATGCCGTTA TCTTTGAAGA 420

ACCTGTTTCC AAGGGATTTG TGGAAAATAA TCCTGATTTA GCAATCGCAG ACCTCAATTT 480

TGAAAAAGAG CAAGATGATT CCTACGCGGT AGCCATgAAA AAAGATAGCA AGAAATTGAA 540

AGAGGCAGTT CGATAAAACC ATTCAAAAGT TGAAGGAGTC TGGGGAATTA GACAAACTCA 600

TTGAGGAAGC CTTATAAGCA TCCA 624

(2) INFORMATION FOR SEQ ID NO: 325:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1237 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 325:

TCTTATGAAG CCGAAGCGTG ATTTATGGCG GATAGGTTTG GTCTGCAGAA AGTGACAAAT 60

CTAGTGCCAT CAGCGTATAT GGAATCTnTG GCTGAGAAAC AGTCCCGGGG TGAACTGACT 120

TATGAGCAGG TTTATGAGGA TGCAACGGCT TATCATCATA CCATTGATGC GAGTACAGAG 180

GAGGCAGACT TGGTTTCTCT ACGTATTGTA GAACTATTGT CTCGAAGAGG CTTTAGCTTC 240

AGTCCTGCGA TCTTACTTGC TATTCATAAG GAGTTGTTTC AAGATATATT TGAACCCTCG 300

ATTCCGGTAG GTCAATTTCG TCAGACTAAT ATCACAAAGA ATGAACCTGT TTTGAATGGT 360

GAAAGTGTTG TGTACTCTGA TTACTCCATG ATTCAAATGA CCTTGGATTA TGATTTTAAT 420

CAGGAAAAAC AAGTTGCATA TGCGACACTA ACCCAGGCGG ATATGGTTAA AAAAATCCAG 480

CATTTTATTT CAGGAATCTG GCAGATTCAT CCATTTCGCG AAGGAAACAC TCGGACGGTA 540

ACGGTATTTT TGATTCAGTA TCTTCGTGAG TTTGGTTTTG ATATTGATAA TACACCATTT 600

CAGCAACATT CCAAGTATTT TCGTGATGCC TTAGTGTTAG ATAATGCAAA GATTTTACAG 660

CGACGTCCTG AGTTTTTAAC AGCTTTTTTT GAAAATCTCT TGCTCGGTGG TCAAAATGAT 720

TTGTCTTCAG AAAAAATGTA TCTAGATTTA GACCTCGATC TTTCATAATC CTAATACTGA 780

GTAAACATTG AATTTTAGGA AAAAATGAAG TAAATATTCT CACAAGAAAA CGTATATCAT 840

CAAAGTTTGG CTCTTTGTCA ATTGTAGTGG GTTGAAGAAA AGCTAAGTTC GAGAAAGGGC 900

AAATTTCGGC CTTTCCTTTT TGATGTTCAG AGCGATAAAA ATCCGGTTTT TTGAAGTTTT 960

CAAAGTTTCG AAAACCAAAG GCATTGCGCT TGATAAGTTT GATGAGATTA TTGGGCGCTT 1020

CCAGTTTGGC ATTAGAATAG TGTAGTTGAA GGGCGTTGAT AACCTTTTCT TTATCTTTGA 1080

GGAAGGGTTT AAAGACAGTC TGAAAAATAG GATGAACCTG CTTAAGATTG TCCTCGATAA 1140

GTTCGAAAAA TTTCTCCGGG TCCTTATTCT GAAAGTGAAA CAGCAAGAGT TTGAAGAGCC 1200

GATAGTGATG TATCAAGTCT TGTGAATAGC TCAAAAG 1237

(2) INFORMATION FOR SEQ ID NO: 326:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 461 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 326:

TTTGATTTTT CTGAATTAGA AGAGATTGAA TTGCCTGCAT CTCTAGAATA TATTGGAACA 60

AGTGCATTTT CTTTTAGTCA AAAATTGAAA AAGCTAACCT TTTCCTCAAG TTCAAAATTA 120

GAATTAATAT CACATGAGGC TTTTGCTAAT TTATCAAATT TAGAGAAACT AACATTACCA 180

AAATCGGTTA AAACATTAGG AAGTAATCTA TTTAGACTCA CTACTAGCTT AAAACATGTT 240

GATGTTGAAG AAGGAAATGA ATCGTTTGCC TCAGTTGATG GTGTTTTGTT TTCAAAAGAT 300

AAAACCCAAT TAATTTATTA TCCAAGTCAA AAAAATGACG AAAGTTATAA AACGCCTAAG 360

GAGACAAAAG AACTTGCATC ATATTCGTTT AATAAAAATT CTTACTTGAA AAAACTCGAA 420

TTGAATGAAG GTTTAGAAAA AATCGGTACT TTTGCATTTG C 461

(2) INFORMATION FOR SEQ ID NO: 327:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1436 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 327:

TAACATTTAG GTACCTCTTC TTAACAAAGT TCAATAGTAA CAATTAATAT TTTAAACAAT 60

ATATCAAACA TCAATGACTA GAATACTTGC ATCATCCTTC TTTCCATAGA TTGGATCAAT 120

AGCAGAAGAA TTAAATCTCA TCTTAATTAA CTCTTCAAAA GTTTTATTTT GATTATTTTG 180

ATAGAATTCA TAAAAGCCAT CGCTCATTAA AACAATTTGT TCACTAGTAA CATCTATTTG 240

ATTAATAATA GCATGGTCTA AAAATCTCTC ATCCAACGAA CCTATCCAGT ACCCACTCGG 300

TTGATTAGAT AATTTTCTGA TTTTTTGTAA AATAATTTTT TTATTTAAAA CACTATTTGT 360

ACCAATTGAA TCTTTTATCT CATTTTTCCC TTTTTCAAAT AAGTTATCTA CTCTATGATC 420

AGTTATTTCC ATTTCGTTTA CTAACATGAC GCAGTCACCT AGCATCATAT ACTCCAACTT 480

TTTTTCTGAA AGTTTAGCAA ATATTGGTAA GCGATAATAT AGTATATTGA AACTAGAATA 540

GTACACCTCT ACTTCTAAAA CATTGTTAGA AATCGATTTG ACTGTCCTGA TTGATTTGTC 600

CTATTATTAT TTCATTTTAC TATACTCTGT TAATTTATAT GAGTTTAAAC CGATTTCATC 660

TTTAACCTCG AGTAAAGCAG TTTCAAATAT TTGTTTAAGA GTTTTTGATT CTTTACAATT 720

AACCGACAAA CTTTCTGATA AAATATGTAC AACTTCTGAG ACTGAATAAC CTATCTCCTC 780

TTTAGAATTA TATAAATCTG TAGCTCCACC AATAATCCAA AAATACTGAT TTTGTGAACC 840

TACAATATCC TCATTTTCTA CGGAACTTCC TTGTATCGAA CAAATTTTAT TTATCTTTAC 900

CATAATACTT CAACCCTTTT AGTGTCAAAA GTAAACCAAT TCCTGTCACT GTTAAGAATA 960

GTTCCATAAT CTTATTCGAA CCAGTCTTTG GTAATTTTTG TTTkACATCT ACTATyTCTT 1020

TAGATTTATT AATATGATTT TCAGTTTCTC TGCCATCTCC AACTATTTTA TAGTTTACTT 1080

CTTCTGTCTT ATTATCTTGT TTATTGTCGA TCTTGTCATT CATTTGTCTA TTATCTTTAC 1140

TTGAGTTAAA CTCTCCGTTC TTCTGGTTAC TATCAATTAC ATTATTTGAA TTAGATTGTT 1200

TTTCCTCTTT GTTTTTTTCT TTTTCGTTTT TATCACTTAA ATTATTTGTT ACAATTTTGT 1260

AAAGCCCATT CTCCGTTACA ATATTGAAAT TACCATCGCT ATCACGTATA ACAGGTTCTT 1320

TCCCATTTGC ATTAGATTTG ATGAATGATA TATACTTACC GGATAAATTA TAAAATTGGT 1380

TATTTAAAAC GGTTATTTTA CCCTTTGAAT CCTCAATAAC AATTCCTTCT TTACCC 1436

(2) INFORMATION FOR SEQ ID NO: 328:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 646 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 328:

CCGGCAGACA GGAGAAGGTG TTAAATATCA ATCTCAAATG GTTCGTCAAT GGTTTCTGAT 60

ACGTATTTTC CGTCTTTCTT CCGTTGCTTG ACACACTCTG TGAGGAGATA TTCGATTTGC 120

CCATTGACTG AACGAAAGTC GTCTTCTGCC CATGATGCGA GTGCAGCGTA TAACTTTGTT 180

GAGAGTCGAA GGGGGATCTG CTTTTTTTTA GCTTCAGCCA TCTTTAGTAA AGGCTTCCTG 240

TGTTGACAAT TGGTTGTGCA TCATGATTGC CACAAAGAAC GACAAGGAGA TTTGAAACCA 300

TGGCAGCTTT TCGTTCTTCG TCAAGTTCTA CCAATTCCCC TTCATTGAGC CGTTCTAGTG 360

CCATTTCAAC CATTCCTACA GCACCATCTA CAATCATCTT CCGTGCATCA ATAATGGCAG 420

ATGCTTGTTG GCGTTGAAGC ATAACGGCAG CAATTTCTGG AGCATAAGCT AGGTAAGTGA 480

TACGTGCTTC AAGGATTTCC AAGCCAGCAT CCTCAACACG ACTTTGGATT TCTTCACGAA 540

TACGGGTAGC AACAATTTCG CTAGAGCCAC GGAGACTACC TTCATCTGCG TGCCCATCAC 600

CCGGAGTATC CACATTAGGA GACACATCGT AAGGATAGAT GCGGAC 646

(2) INFORMATION FOR SEQ ID NO: 329:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1653 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 329:

GTTGCAGGTG CAGTAGGTGT TACTTCAGAT ACATTTGAAC GTGCAGAGGC TCTTTTTGAG 60

GCAGGAGCGG ATGCGATTGT TATTGATACT GCACATGGTC ATTCTGCAGG TGTCTTGCGT 120

AAAATTGCCG AGATTCGTGC TCATTTCCCA GATCGGACTT TGATTGCTGG AAATATTGCT 180

ACTGCTGAAG GTGCACGTGC CCTTTATGAA GCGGGTGTAG ACGTTGTTAA GGTTGGTATT 240

GGACCAGGTT CTATCTGTAC TACTCGTGTG ATTGCTGGTG TTGGTGTTCC GCAAGTAACA 300

GCTATCTACG ATGCTGCAGC TGTTGCGCGC GAATATGGTA AAACGATTAT TGCTGACGGT 360

GGGATCAAGT ATTCTGGAGA TATTGTAAAA GCACTTGCTG CAGGTGGAAA TGCTGTTATG 420

CTTGGATCTA TGTTTGCTGG AACTGATGAA GCTCCAGGCG AAACTGAAAT CTTCCAAGGA 480

CGTAAATTCA AGACTTACCG TGGTATGGGA TCAATTGCTG CTATGAAGAA AGGTTCAAGC 540

GACCGTTATT TCCAAGGTTC TGTCAATGAA GCAAACAAGC TTGTTCCAGA AGGAATTGAA 600

GGTCGTGTTG CTTATAAAGG AGCGGCAGCT GATATTGTTT TCCAAATGAT TGGTGGTATT 660

CGCTCTGGTA TGGGTTACTG TGGTGCAGCT AACCTTAAAG AACTACACGA TAATGCTCAA 720

TTTATTGAAA TGTCTGGTGC TGGTTTGAAA GAAAGCCATC CTCATGATGT GCAAATTACT 780

AATGAGGCAC CAAATTATTC TATGTAAAAA ACAATGAAAA GAACTCCAGT GAAAACAGGA 840

GTTCTTTTAC AATGTTGTCA ATTTCCATTT ACAGCAGCTT TACCATCCTG AATAGTGAAG 900

ATACTTAGAT TTTCTGGCAG ATTTTGAAGA TGGTCTAAGC TTGTTGTTGT GATAAAGGTT 960

TGGATTGATT GAGAAATCGT TTCTAATAAT TTTAACTGTC TAGTGTTGTC AAGTTCACTC 1020

ATCACATCGT CAAGCAGTAA TATAGGAGAT TCTGTGGTAA TGCTTTCCAT TAATTCGATT 1080

TCTGCTAATT TTATCGAGAG GACGAGACTA CGATGTTGAC CTTGGCTTCC GAAACTAGCA 1140

TCCATCCCAT TTATATAAAA AGAAATGTCA TCTCGATGAG GACCGACACC AGTATTCTTT 1200

TTAAATAAAT CTCTGGATCT ACTTTTTTCT AAAGCAATTT TGAAAGATTC GGATAAGTTT 1260

TGTTTGTCAG TTATATTGAC AGAAGATTGA TAGGATATTG ACAACTCTTC GATCTGATTA 1320

GAGAGTTCAA AATGTTTCTT ACGCCCAAAT GATTCTAGTT TTTTTATGAA ATCTAAGCGG 1380

TGATTCATTA CACGACATCC ATAATCAACT AGCTGATCAT CTAACACAGA AAGGAATGTT 1440

TCATCTATTT TTTGAGCTGA TTTTAGGTAA GTGTTTCTTT GCTTTAGGAT GTGGTTATAA 1500

TTGGTTAAGT CAGATAAATA GATTGGCTTA ATTTGCCCAA GTTCCATATC AATGAATTTT 1560

CGTCGAATCG AAGGTGCTCC TTTAATTAGT TGTAAATCTT CAGGAGCAAA TAAGACAACA 1620

TTCATGTGTC CTACATAATC TGAAAGGCGT GCC 1653

(2) INFORMATION FOR SEQ ID NO: 330:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1340 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 330:

GAAACACTGT ATTTCAAAGC ATTTTTTGTT AGTTTAAAAT TACTCCCATT CTTCTTTTCC 60

AAACGTACAA TATATCCAAA ACCATTCAAA ATACTAGATT CTATTTTTTA TAATATCACT 120

AAATCCACCT AATTATAGGA CGTTTTCAGA TTTTTAGTCC CAGTCCCAGT ACCGGAGAAA 180

TATTGTTTTA ATATAATATC TCTTTTTGTC TTCTAAGCTC TTAAAAGCAA AAGAACAAGT 240

AAAGAGTCAA GACAAGGATA AAAAGTCCAT ATTAGGGCAA ATAAAAAGCT TTAAGACAGA 300

TGACAAATCT AAGTCAAATA AGAAAGACCA TAGCAAAGGT GCAGAGAGAT AAATATTGGC 360

GGTCTTCGGA CTGCCTTTAT TTTTTTATCC ATTTTTCAAA TCAAATTTAT TCAGACTATA 420

TATGCACATA TACACTTAAA TTCATATAAA AACATGGCTT GTAAAAAATT ACTTTAATCA 480

CAATAATCGC ATTTAAAATT GTGATGTTTG CAAGCTAAAT TACGGACTTC ACTTGGAAGT 540

TTTCCCTTGT ATCTTTTATA ATAGATAGAA AATTTGCTGG CAGATGAATA TCCAACAGAT 600

TCTGCTATCT CTTTTATAGG TAGTTCAGTG TTTAAAAGAA GAGTTTCAGC TACATTCATT 660

CTTTTTCTTT GAGTGTACTC TGTAATGCTT TGACAATATT TTTCCTTAAA TAAATTTTTT 720

AATTTAGTAC CACTCATTTT AGATATTTTT TCAAGCGTGC CTTGATTTAC ATTCGTTGCA 780

AAATGATCAT CTAAGAATCT TGCTACATCT TCAAGTGCTT TATCATCATC AATTTCAATT 840

TTATATTTTT TTCTATTTAA GTATGTGTCA ATTACTATAC TTATCCATTC ATTTGCCTTT 900

GCTTTAAAGA AAAAATCAGC GGCAGGAGCG TCCATCTTAC AATTTAATAT TTCCATTGCC 960

ACTCTTTCTA AGGCCTTTGT AAGTATTATT TGATTCGGTT GAAGCAAGGT TGAATAAAAA 1020

GATTCTGGAT TAATGTTAAT AGATGCTAAA TGTTTTTCTA TTAGCTCTTT TTTAAAACCm 1080

ATGGAAACAG CAAGATAACA ACAATTCTCG TGTAATAAAA AAACAAAATT ATCTTTTATA 1140

TTATCAAAAT CAAAAGTACA TAGAGAGTTT GCGGTAATAG TTTGATACGG ATTAAACTTT 1200

TCTCCGTTTG CACTGACAAT GTAACTTGAA TAAATTGAAA CATAGTCTGA CATACTATAA 1260

GTGCTATTTT GAACTACTTC CTCTTTGATA TAAAAATCAT GTATATCGAT AATGAAGATG 1320

CCTCCTTCAT AAAACCGGTA 1340

(2) INFORMATION FOR SEQ ID NO: 331:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 607 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 331:

TATGTTCGTG ATGAGTTTTT AAGTAGGAAA AACGTGCTAA CCTCTCAGAT TTTGGAACTT 60

GTAAAAGAAA CTCTTTTTTC ACCCGTAGTA GTTGATAATG GGTTTGATCC GGCCTTATTT 120

GAAATTGAGA AAAAACAATT GCTAGCAAGT TTAGCAGCTG ATATGGATGA TTCTTTTTAT 180

TTTGCACATA AAGAATTGGA TAAATTGTTT TTTCATGATG AACGTCTTCA ATTGGAATAT 240

AGTGATTTAC GAAATCGTAT TTTAGCTGAA ACTCCACAAA GTTCTTATTC TTGTTTCCAA 300

GAATTTTTAG CCAATGATCG AATAGATTTC TTTTTCCTAG GTGATTTTAA TGAGGTTGAA 360

ATTCAAAATG TATTAGAATC ATTTGGCTTT AAAGGTCGAA AAGGAGATGT GAAGGTTCAG 420

TATTGTCAAC CTTATTCTAA TATCCTTCAG GAAGGTATGG TTCGGAAAAA TGTGGGACAA 480

TCCATTTTGG AATTAGGTTA TCATTACTGT TCTAAATATG GTGATGAGCA ACATTTACCC 540

ATGGATTGAA TGAATGGTTT ACTTGGTGGA TTTGCTCACT CTAAGCTCTT TACAAATGTC 600

CGGGAAA 607

(2) INFORMATION FOR SEQ ID NO: 332:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 900 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 332:

TTAAAATACC GAATTTTGTT TTGTCCTCTA TTTCAACATT GTGAATCGCC TCAGGCAGAG 60

AACCGATACT AAAGATATAA CCAAAATAGT TGTCATTTGC TTTACCGATA TCAATCTTAT 120

TGGTTAAATC AAAATCCAGT TCGTCAATTG CGCCATCGAT GTCTTGATTG ATTTCCAAAA 180

GTTTTGTAAT GAGGTTACCC GTACCGCCTG GGATAATCCC TAACTTAGGA ATGTAGTCTC 240

TCTCATCAAT ACCTGAAATG ACTTCATTGA CAGTTCCATC TCCACCAAAC ACAACCACTG 300

CATCATACTG CTCACGAGAA GCTTCTTCAG CAAAATGTGT TGCATCCAGC GCTTTTTCGG 360

TAATTTTGGT TTCAACATAT TCAAAGTATT CTTTTGCTTT ATTCTCCAGC TTTTCTTTGT 420

AATCCAAAGC CTTCTCGCCA CCAGAAGTAG GGTTGATAAT TACCATTGCT TTTTTCATTG 480

ATTTTATCCT TAATTTTAAA CAGAAATGTT TACATTTCGT CGTATGCAAG TAAATGTAAT 540

CCTATTATAC AATGAAAATA CAGAAAAGAG AAATCTGACG TACTGGAGAT TAATACGCTT 600

TTATTCTATT TTCCCATCGC CTAACTACAT CCTTTAAGGG TTCATCCAAG TAAGAATAGG 660

CCTTATCCTT GATCCAATCA GGAATACCGT AAGCTGCCTC TGCTAwGCTA CAAGTGATTG 720

CTGCGAGAGT ATCACTGTCG CCACCAAGTG AGATGGCATT TCTTATCGCA TCTTCGAAGT 780

CTCTACTTTC AAGAAAGGCG ATAATGGCTT GAGGGACAGT TTCCTGACAT GTTTCGTTAA 840

AACGATAGTT AGGACGGATT TCATCTAAAG TTTGAGATAG ATTGTAATCG TATTCTTTTT 900

(2) INFORMATION FOR SEQ ID NO: 333:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 533 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 333:

CCTTTCTGGC ACACTGGTCT TGGAATACGG CAAAACCTCT GAAAATATCT ATGCTGGAAT 60

GGACGAGGAA TACCGTCGTT ATCAGCCTGC CATCATCACT TGGTACGAAA CAGCCAAACA 120

TGCTTTTGAT CGCGGACAGA TTGGCAAAAT ATGGGTGGAA TCGAAAACGA CCTCAAGGGC 180

GGTCTCTACA GCTTTAAATC CAAGTTCAAT CCGACCATTG AGGAATTCGC TGGTGAGTTC 240

AACCTGCCAA CTAATCCTCT TTACCACCTC TCCAATCTGG CCTACACTCT CAGAAAGAAA 300

CTGCGCAGaA GcATTAACAG AAAGGAAGCC TATGACCTTT AAACTTCTCA GCCAAGAAGA 360

ATTCATCCAG CATACCTCAG CTAGATCCCA ACGCTCTTTT ATGCAGACCG TAGAAATGGC 420

AGAGCTGCTG AGCAAGCGTG GCTTCAGTAC CCAGTATGTC GGCTACACTG ACCCACAAGG 480

GAAGGTAGTG GTGTCAGCTG TCCTCTACAG CATGCCTATG ACTGGTGGCC TTC 533

(2) INFORMATION FOR SEQ ID NO: 334:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 544 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 334:

CCAGCAAACT AGGAAGCTAG CCGTAGTTGC TCAAAGCACA GCTTTGAGGT TGTAGATAAG 60

ACTGACGAAG TCATGTACAA AACACTGTTT TGAGGTTGCA GATAGAACTG ACGAAGTCAC 120

TCAAAACACT GTTTTGAGGT TGCAGATAGA ACTGACGAAG TCACTCAAAA CACTGTTTTG 180

AGGTTGCAGA TAGAACTGAC GAAGTCAnnA ACCACACCTA CGGCAAAGTG AATCTGAAGT 240

GGTTTGAAGA GAGTACAACT TGTCTTTTAG AAAAGGAGCC TATAATGAAA GTCTTTCAGC 300

ATGTAAATAT CGTGACTTGT GATCAAGATT TCCATGTTTA TCTTGATGGA ATCTTAGCAG 360

TCAAGGATTC TCAAATCGTC TATGTCGGTC AAGATAAGCC AGCGTTTTTA GAGCAAGCTG 420

AGCAGATTAT AGACTATCAG GGAGCTTGGA TTATGCCTGG TTTGGTCAAT TGTCACACCC 480

ATTCTGCAAT GACAGGTCTG AGAGGGATCC GAGATGACAG CAATCTCCAT GAATGGCTCA 540

ATGA 544

(2) INFORMATION FOR SEQ ID NO: 335:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 349 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 335:

CCAGGAACTC AAATGTAAGT AGGGGTTCCT TTTTTGTATA TTTTTCAAAT AACGCCTCTA 60

CACTATTTGT AGCAAATTCA CCAACTACAG TTGTATCTTA GTTAAAATAA GTTAGAATAT 120

GTAAGTGAGT ACCAGATATA CCAAGACATC GTCACCATCT AAGGTATATT CAAAATACAA 180

AAGTTGACCA ACTAGATTTC TGAATATCCT TATATATCCA TTCTTAAAAT TGGTTTAAAT 240

AGCGTAGTCT TTTAAACTAG TTTTGAGAAT CCAAAAAATC TTCCTACATA TGTAAGAAGA 300

TTTTTTAGTT CAGAATGATT AGaTTTAGCT AATGGATACC TATCCTACC 349

(2) INFORMATION FOR SEQ ID NO: 336:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1206 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 336:

CTCCGATAAC CACACCAGCA ATGGAAATAA TTCCATCGTT AGCATCAAGA ACACCCGCAC 60

GCAGGATATT TAAACGACCT GCAAAATTTG AATCAATTTC GTGATTTGTT TCTGACGCTA 120

AATTTCAAGT TCAAGTTAGC CATCAAGAAG TCTTCTCTGG GTGACTTGTA GTCCAAGCAT 180

TTTTTAGGAT AGTTGTTAAT CCACTTTTCG ATGAATGCGA CTTCTTTGGG AGTCATTTTC 240

TTGGTTCCCT TAGGTAACCA TCTACGAATG AGCCTGTTGT GATTCTCATT AGTTCCCGGG 300

ATCCTCTAGA GTCGACCTGC AGGCATGCAA GCTTGGCACT GGCCGTCGTT TTACAACGTC 360

GATGACTGGG GAAAACCCTG GCGTTACCCA ACTTAATCGC CTTGCAGCAC ATCCCCCTTT 420

CGCCAGCTGG CGTAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC AGTTGCGCAG 480

CCTGAATGGC GAATGGGGCC TGATGCGGTA TTTTCTCCTT ACGCATCTGT GCGGTATTTC 540

ACACCGCATA TGGTGCACTC TCAGTACAAT CTGCTCTGAT GCCGCATAGT TAAGCCAGCC 600

CCGACACCCG CCAACACCCG CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCCGC 660

TTACAGACAA GCTGTGACCG TCTCCGGGAG CTGCATGTGT CAGAAGTTTT CACCGTCATC 720

ACCGAAACGC GCGAAACGAA AGGGCCTCGT GATACGCCTA TTTTTATAGG TTAATGTCAT 780

GATAAGGATG GTTTCTTAGA CGTCAAGTGG CACTTATCGG GGAAATGTGC GCCGAGACCC 840

TATTTGTTTA TTTGTCTAAA TACATTCAAA TATGTATCCG CTCGTGAGAA AATAAACCTG 900

ATAAATGCGT CAATAATATT GAAAAATGAA GAGTATGAGT ATTCTACATT TCCGTGTCGC 960

CCTTATACCC TTTTTTGCGG CATGTTGCCT TCCTGTTTTT GCTCACCCAG AAAACGCTGG 1020

TGAAAGTTTA AGATGCTGAA AAATCATTTG GGTGCACAAC TGGGGTTACA TCCAACTGGA 1080

ATCTCCAnCA GCAGTTAAGA TCCTCTGACA GTTGTACACG CCGCAAGAAC TATTCCCGAT 1140

GAATGAGCAA CTTTTAAAAG TCCTGCGAAT GTTGGGGCGG TAATAATCCC CGTGTTGTAG 1200

GCCCGG 1206

(2) INFORMATION FOR SEQ ID NO: 337:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 813 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 337:

CTGCTCAACT CAGACAGTCA AATTTCTGAC TTTACCAAAA GAACCATCAA AAAAGTTGCT 60

GAAAAAGGCC ATCAGGTTAT TATTACGACA GGTCGCCCTT ACCGTATGTC AAAAGATTTT 120

TACCGTGAAC TGGGCTTAGA CACTCCTATG ATTAACTTCA ACGGATCCCT TACTCATTTA 180

CCAGACCAAG TTTGGGATTT TGAAAAGTGT TTGACTGTAG ACAAAAAAATA TCTGCTAGAT 240

ATGGTTCAAC GTTCAGAGGA CATTCAAGCC GATTTTATCG CTGGAGAATA TCGTAAAAAA 300

TTCTACATTA CAAATCCCAA TGAAGAAATT GCCAATCCCA AACTATTTGG TGTAGAAGCT 360

TTCCAGCCTG AAGATCAATT CCAGCCTGAA TTGGTGACCA AGGACCCTAA CTGTATCCTC 420

TTGCAGACTA GAGCCAGTGA CAAATATTCC TTGGCAAAAG AAATGAACGC CTTCTACCAG 480

CATCAACTTT CTATCAATAC CTGGGGAGGT CCGCTCAATA TCCTTGAATG TACCCCAAAA 540

GGTGTCAACA AGGCCTTTGC TTTGGACTAC TTGCTCAAGA TAATGAATCG TGACAAAAAA 600

GATTTGATTG CCTTTGGAGA TGAACACAAT GATACCGAAA TGCTCGCTTT TGCTGGGAAG 660

GGTTATGCCA TGAAAAATGC CAATCCAGAG CTACTCCCTT ATGCAGATGA GCAAATTTCC 720

CTTACCAACG ACCAAGATGG GGTTGCCAAA ACCCTACAAG ACTTATTCTT ATAACCTATA 780

CTGATACTCA ATGAGGGGCA AAGAGCGAAC TTA 813

(2) INFORMATION FOR SEQ ID NO: 338:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 683 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 338:

CCTAGATAAA TGATATAATT CTATTATTGT TCGTAAAAAT TAAAAGGAGA TTGATGATGG 60

1360

ACAAATTATT TAAACTAAAA GAGAACGGTA CAGACGTTCG TACAGAGGTT CTCGCTGGTT 120

TAACAACTTT CTTTGCAATG AGCTATATTC TCTTTGTAAA CCCACAAATA CTTTCACAAA 180

CAGGAATGCC TGCTCAGGGC GTCTTCCTAG CGACGATTAT TGGTGCAGTA GCGGGTACCT 240

TGATGATGGC TTTTTATGCT AACTTACCTT ATGCCCAAGC GCCAGGTATG GGACTCAATG 300

CCTTCTTTAC CTTTACAGTT GTATTCGGGC TTGGTTATTC TTGGCAAGAA GCCCTAGCTA 360

TGGTCTTCAT CTGTGGGATT ATTTCATTGA TTATTACCTT GACAAATGTT CGTAAAATGA 420

TCATTGAATC GATTCCCAAT GCTCTTCGCT CAGCTATTTC AGCTGGTATC GGTGTCTTCC 480

TTGCCTATGT AGGGATTAAG AATGCTGGAC TTTTGAAATT CACGATTGAT CCAGGCAACT 540

ATACTGTTGT AGGAGAAGGG GCTGACAAAG CTCAAGCAAC GATTGCAGCA AACTCTTCAG 600

CAGTTCCAGG ATTGGTCAGC TTTAATAATC CAGCTGTTTT AGTGGCTCTT GCAGGACTTG 660

CCATTACTAT CTTCTTTGTC ATC 683

(2) INFORMATION FOR SEQ ID NO: 339:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 852 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 339:

CTACTTTACA TGGAAGTAGT CACTGAATTC CAGTTAGAAA TTACTTTGTA ACTACGTTTT 60

GAGGAGGAGT AAAATGCTTT CCTACGTTCG ATATTACCCA CTAGCGATAG CTAAATTAAT 120

GTGTCTGTGC TCTCCTAAAA TCTGCTGATT TATTACTGAC TAATACAGGA GGTTTTTTTT 180

ATGgACAGAC AATCATATCT GCTATTGGTG TTTATATTTC CACCAGTATC GATTATTTAA 240

TTATTTTAAT TATTTTATTT GCACAGCTAT CACAGAATAA ACAGAAATGG CATATTTATG 300

CGGGGCAATA TCTAGGCACA GGCTTACTTG TAGGGGCGAG TTTAGTTGCT GCTTATGTCG 360

TTAATTTCGT GCCTGAAGAA TGGATGGTTG GATTGCTTGG TTTAATCCCT ATCTATTTAG 420

GGATTCGCTT TGCAATTGTT GGAGAAGATG CGGAAGAAGA AGAGGAAGAA ATTATTGAAA 480

GATTAGAACA AAGCAAGGCA AATCAACTGT TTTGGACAGT TACATTGCTG ACAATTGCGT 540

CTGGCGGAGA TAATTTAGGT ATCTATATAC CTTATTTTGC TTCGTTAGAT TGGTCACAGA 600

CCCTCGTGGC CTTGCTTGTG TTTGTAATCG GCATAATTAT CTTTTGCGAG ATTAGTCGGG 660

TGTTATCCTC TATTCCGTTA ATATTCGAGA CAATTGAAAA ATACGAGCGA ATCATTGTGC 720

CCTTAGTATT CATTCTACTT GGACTATACA TCATGTATGA AAATGGCACG ATAGAGACTT 780

TTCTGATCGT GTAGATTTTT TTGTTTCACT AGGGATTTAG CCCGAGCTCA AATCAGCTCT 840

CTGATTTTCA GA 852

(2) INFORMATION FOR SEQ ID NO: 340:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 754 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 340:

CCGCACAAAA GCGCATAGTA TCAAGATTCT ATAAAGCCTT GATACTATGC CTTTTTAATG 60

GATAAATAGT TAGTCTTTTT TAAAGACCGG ATCTTTCAAA CTCTGCATAC TGGCATTGAT 120

CACCGCGCCT AGGATAACAA TTTTAGCAAT CAAGATAAAC CAAAACATCA TAACAACAAG 180

AAGAACGGAA CCTAAAATTC GGACATCCAC CAAATGATGG ACATAGTAAT TGAGATAACT 240

AGAGAACAGA GTTAGTAAAC CTAAAATCAC TAAGAGAACA AAGGCACTGC CTGGTAGGGT 300

ATAGCTAATT TTCCTGTTAG ATAGATTGGG AAGAAAATAA TAAAGCATGA CCAAGATAGC 360

AAAGAGGAGG GCGTAAATCA GAGGACCTGC CAACCCTTGT AAAGCCTGAT AGATAATGCC 420

ATCTTTTGTC CAATAATGAG CAAGTAAAGC CAAAATCATC TGACCAAATA AGATCAAAAA 480

CAAGGCAAAC GCAAAGAGGA GCTGCAACCA AAACTGACTA GGAGACTTAG CATCTGATGG 540

GAAATAAGTC CACGACTCTT TTCGACGCCA TAAGCCTTGT TAAAAGCTTT TTGCAAGAAA 600

TTCATAGATT TTGAAAAACT CCATAACGCC GATAAAACAG AAAAACTCAA TAAACCTGTT 660

GAAGGTTGCG TCAAGACTTC TCTGGCTATT TTTTCCACAC CTTCATAGAG GCTTGGGGGG 720

CAGACGTCTT TCATAAAGCC CAAAAATTCT CCCA 754

(2) INFORMATION FOR SEQ ID NO: 341:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 707 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 341:

GGGGATAACT CTAGGAGTAC CGCTATTACT CGACTTAATG AGTGCACAAG AAGTCAGGAT 60

TTTTATGCAG GTTGGGCGCT TCATCAGACA GGGAAGATTT ACAGCGACTA TTATGGAAGT 120

CAAGGTTTGC TTTATTATTT GCTGACTTAC GTGAGTCAGG GCGGATTTTT CTTTGCCATC 180

TTTGAGTGGT TAGCCTTGGT AGCAGGAGGA TTTTTCCTTT TTAGATCAGC GGACACCTTG 240

ACAGAGCAAG GAGACCAAGC TGGACATCTG GTGACTATTT TTTACATGCT AGTTACAGGT 300

CTTGCTTTTG GTGGAGGCTA TGCGACTCTT TTAGCGCTTC CTTTCTTATT CGCAGCCTTT 360

AGTTTAGTTG CGGCTTACCT AAGCAATCCA AGCCATGATA AGGGATTTGT ACGGATTGGG 420

CTAGCTTTGG CAGGCGGATT TTTCTTTGCT CCCTTATCAT CGCTCCTGTT TATTGCTGTA 480

GTGAGTTTAG GCTTGTTGGT CTTTAACCTT GGGCATAGAC GCTTTGCGCA TGGGTTTTAT 540

CAGTTTCTTG CAGTGGCTTT AGGTTTTTCA CTTGTCTTTT ATCCAACTGC CTACTATAGT 600

GCTGCAACAG GAAGTTTTGG GGATGCGwTT AGTGGTATTC GTTATCCTAT TGACAGTATT 660

CGCTTTGATT TTACTTCTAA AATTTTAGAG AATATGTTTT TTTAAGG 707

(2) INFORMATION FOR SEQ ID NO: 342:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 762 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 342:

GGATTTTGAA AAACCATACC GATTTGACGA CGTATATTCC AAACATTTTC CTCAGTCAAA 60

CGTTGGCCAT CAATTACAAT CTCTCCGGAT TCTGCTTCCA GTAAGCCATC AATTAATCGA 120

ACCGTCGTTG ATTTACCACT ACCATTATGC CCTACAATCG AAAGCCATTC TCCACGTTTC 180

ACGTGAAAgT AATATCCTTC ACATCGTAGT AGTTCTGATT TTCTTTATAG CGAAAAGAAA 240

GATTTTTTAC ATCAATTATT GATTTCATTT CGAACCAAAT GTCCCTTTAA ATACATAGGC 300

ACTACCCTTG AAATAGTCAT AGCCAGAGTA GATAGTGAAA AATAAGGCTA CATAAAGTAG 360

AACTTGACCA AGCAAAGTCC AATGTAATAG CAAGAAAATA ATGGCAAACA TCTGACTAAA 420

AGTTTTAATT TTTCCAGGCA TTGCTGCTGC TAAAATTGTT CCACCAGTTT CAACCAATAA 480

AAGCCTTAAA CCTGTCACAG CTAACTCACG ACAGATAATC ACTGCAACAA TCCAAGCCGG 540

AGCCATACCT AACTCAATCA ACATAATAAA AGCCGACATA ACTAGTAACT TATCCGCCAT 600

AGGATCTGCA AATTTACCAA AATTACTGAC CACATTCCAT TTACGAGCTA AATATCCATC 660

TAAATAGTCG GTAATACTGG CAACAGCAAA GATAATAGCT GCAACTATAT GACTCTCTAT 720

CGAATTTCCT ATCGTTAAAA TAAAGATAAA AATAGGTATA AA 762

(2) INFORMATION FOR SEQ ID NO: 343:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 482 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 343:

CTTTTGATAC ACTTAAACTA TGAATACAAA TCTCAAGCCC AAACTTCAGC GTTTTGCTTC 60

TGCGACTGCC TTTGCCTGTC CTATCTGTCA AGAAAATCTG ACTCTGTTAG AGACTAATTT 120

CAAGTGCTGC AACCGTCATT CTTTTGACTT GGCGAAATTT GGCTATGTCA ATCTAGTCCC 180

TCAAATCAAG CAATCTGCTA ACTACGACAA GGAAAATTTT CAAAACCGTC AACAAATCCT 240

AGAAGCCGGC TTTTACCAAG CTATCTTAGA TGCTGTATCT GACTTGCTTG CAAGCTCAAA 300

AACTACCACA ACAATTTTGG ATATCGGTTG TGGTGAAGGA TTCTATTCTC GCAAACTACA 360

AGAAAGTCAC TCTGAAAAAA CTTTCTATGC CTTTGACATC TCCAAAGATT CAGTCCAAAT 420

CGCGGCTAAA AGTGAACCCA ACTGGGCAGT CAATTGGTTC GTTGGCGACT TGGCACGACT 480

TC 482

(2) INFORMATION FOR SEQ ID NO: 344:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 520 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 344:

TTTATTTTTA TAAAGTCAAT ACCTGTCTTT ACTTTTTCTT AAAAAAAGTT TATTATGTTC 60

TTTAAGGAGG TGTAAAACAT GAAAATAAAT AATAAACTCG TTGGAGAACG TATTCAAAAT 120

ATCCGTTTAA GCCATGGCGA CTCTATGGAA AAATTTGGAG AAAAATTTAA TACTAGCAAA 180

GGTACAGTTA ACAACTGGGA AAAAGGTCGC AATTTACCAA ATAAAGAAAA CCTACTAAAA 240

ATTGCATCTA TTGGAAAAAT GAGTGTTGAA GAGTTACTCT ACGGCGATTA CAATACTTAT 300

CTACACTTAA AGATTATGGA TTTAGCTCCT GAATGTATAA AAAATTATGA TGAGTATAAC 360

TCTTTACACG ATGATATAAC AAATAAAGCG TTACAGATCG CTCAAAATAC CATTTCTAAG 420

ATTGATTATC AAATTTCAGA CGAAACGATC AAAAAATTTA TTGATTTAGC TATCGAACAA 480

TCGAGAGATT TGCAAGGAAA TTTGTTGAAA AATAACGGGT 520

(2) INFORMATION FOR SEQ ID NO: 345:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1003 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 345:

GCATCAAATC CGCCATCAAA GAAGTTCTCT GGATTTACCA AGACCAGTCA AATAGCTTAG 60

AAGTGCTTAA TGACAAGTAC AATGTTCACT ACTGGAATGA CTGGGAAGTT GGAGACACGG 120

GAACCATTGG TGAGCGCTAT GGTGCCGTTG TTAAGAAACA CGACATTATC AATAAGCTTC 180

TCAAACAGTT GGAAACCAAT CCTTGGAACC GCCGCAATAT TATTTCGCTC TGGGATTACC 240

AAGCTTTCGA AGAAACAGAT GGGCTGCTCC CGTGCGCCTT TCAGACCATG TTTGATGTTC 300

GGCGTGTTGA TGGGGAAATC TATCTGGATG CGACCTTGAC CCAGCGCTCC AATGATATGC 360

TGGTGGCCCA CCACATCAAC GCTATGCAGT ATGTGGCTTT GCAGATGATG ATTGCCAAAC 420

ATTTTGGCTG GAAGGTTGGG AAGTTCTTCT ACTTCATCAA CAACCTCCAT ATCTATGATA 480

ATCAATTTGA ACAAGCTCAG GAATTGCTCC GTCGGGAgCC GTCAAACTGC CAACCACGCT 540

TGGTTTTAAA TGTTCCTGAT GGGACTAATT TCTTTGATAT CAAAGCAGAA GATTTTGAGT 600

TGGTGGATTA TGACCCTGTT AAGCCACAGT TGAAGTTTGA CCTAGCTATT TAAAAGAATA 660

GAAAAAAGAA GTTGAGAATA ATCCCAACTT CTTTTGTTTC TTAACGTGAT ACGCGGCGAC 720

GAGCTGCTTT TTTACGGTTT TCTTCGATGA AAGCTGCTTT TTGCTCTTCT GGTTCGATTA 780

CTTTCTTTTT AAATGCGTAT ACTGCACCTG CAACGGCAGC GACAGTTCCT GCGACACCTG 840

TTACAAGACC TTTAGCGAAT CCTTTAGCCA TGAGTCTTCC TCCTTTATAT TCTCAATCAG 900

CCAGCCTCCT CAAGAGGTCA CATTTTTCTG ACTGACCTTT TTGTGTTATA ATAATAGTAA 960

CGAAAAAATG GGAATTTTTC AAGGAAAAAA GATGAGAACA AAA 1003

(2) INFORMATION FOR SEQ ID NO: 346:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 750 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 346:

CCGCACGTAC TATTCCAGAT GCCGAGGAAG TGGACCTCAT CCTCGTTGGC GCAACTGGTC 60

TCAACGCCTT TGAACGCCTC TTGGTCGGCT CTTCATCTGA ATACATACTC CGCCATGCTA 120

AGGTCGATTT GCTGGTTGTG AGAGAACAAG AAAAAACCTT ATAATCACAA AGAAAAGGAG 180

CCCCTAGCTC CTTTTTGTTT ACGATTTATT TCTCTCTTTA TGGCGTTCGT AAGCCTTGAG 240

CTGGCGCTGC AGTTCCTTTT TAATAGCAGG TTCTGGAGCA TATTTTTCTT CCCAATTATC 300

TGGTTTTAAG ATTTTATGGG TCACTGGATC AAAATGAGCC TTGCCATCTG GAAAAATTTT 360

CCCCATATTG GCCTGATGGA CAATATCAAA AATACGTTCT GGGTCCACCC CCATCAAGAC 420

AAAACTGCCG TAGGTGAAGT AAAGCGTGTC AATCAAGGCA TCCACTTGCC CTATCAAATC 480

TTGCTGAGCA GGTGTCTTCT TGGCTACTTT ATCTGCTGCC TTATCAAGGG CCTGATGAAG 540

TTGCGATACA GCTTGACCAA AATCTTCTTC AGAAGGACTG GCTGCTCGAA CAAACTCCAC 600

CAATTCTTCT ATTTTAAAAC CAGCCCTATG GGTTGCACCC TCTAAATCCC AAGCTCGAGG 660

TTCTTCTTGG GTTCGTTCAT CCATCATGTG GTGGAAAGTC TTGACCTTAT TGAAATGATA 720

GTCACGGCTG ACAAAGACTT TTTCTGAAGA 750

(2) INFORMATION FOR SEQ ID NO: 347:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 596 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 347:

CGCAACATAC GGATAACCTC CAAAGAATAT TTTTATATTA TAGCAAAGCT TTAAATTGAA 60

TGTTAGAGTC TTGTTCAAAA CAATCATCAA AACCACGTGG ATGATGGTAT TCTACTAAGT 120

GTTGATCTTG AGGATAAGTG TACTTACCGC CAACTTCCCA GATAAATGGA TGGAAATCGT 180

ATTGCAAGCG ATCTTTTCGC ATTTTCCAAA GTTCTAGAAT CTCATTAGTA GAAGCCATGA 240

AGTTAGACCA GATATCATAG TGAACTGGGA TAATGACTTT GGTACGCAGA TTTTCTGCCA 300

TACGAAGAAG GTCGATAGAT GTCAkTTTGT CTTGGATACC TACCGGATTT TCACCATAGT 360

TATTCAAAGC AACATCAATT TTAAAGTCTT TACCATGTTT TGCAAAATAG TTTGAGAAGT 420

GAGAATCTGC ACCATGATAG ATGGTTCCAC CTGGTGTTTC AAAGATATAG TTAACAGCCT 480

TTTGAGCCAT TTCTTCATCT GTAACAGCCA AGCCAGCAgT TCACCGCCTG TCTCATCAGC 540

ACCGTTCACT GGGAGAGTTA CCAAGCAAGT ACGGTCAAAT GATTCTACTG CATGAA 596

(2) INFORMATION FOR SEQ ID NO: 348:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 673 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 348:

CAGAGTCAAC AGCCTGAGTT GAAGGCAACT TTAGACACAG CAGTTACGAC AGCTGAATGA 60

GCTCCTCCAT CAGTTTTTTC TTTAATGAGT CCAGCTACAT CTTCAACTTC GAGGCCGTTA 120

ATCACAATGT CAGCGCCTAC TTCTTTTGCA AGGGCAAGTT TGTCATTGTT GATATCGACT 180

GCGATAACAT GAGCATTGAA TACTTTTTTA GCGTATTGAA CAGCGAGGTT ACCAAGTCCA 240

CCAGCACCGT AAAGAACAAC CCATTGGCCT GGTTCAACTT TTGCTTCTTT GATAGCTTTA 300

TAGGTTGTTA CTCCAGCACA TGTGATAGAA GAAGCTTGGG CTGGATCAAG TCCGTCAGGA 360

ACTTTGACAG CATAGTCAGC AGTTACGATA CATTGTTCAG CCATACCACC GTCTACTGAG 420

TAGCCAGCAT TTTTCACTGT ACGGCAAAGG GTTTCGCGAC CAGTTGTACA GTATTCGCAA 480

GTGCCACATC CTTCAAAGAA CCAAGCAACG CTGACGCGGT CACCGACTTT AAGGCTTTTC 540

ACATCTGGAG CAATCTCTTT AACGATACCG ATACCTTCGT GCCCAAGAAC ACGTCCTGGG 600

ACTTGACCAA AGTCACCATG AGCAACGTGG AGGTCGGTGT GGCAAACGCC CACAGTATTC 660

ACTTCTACAA GTG 673

(2) INFORMATION FOR SEQ ID NO: 349:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 198 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 349:

GTACCCTACA AATGCTTTAC AGTATGGGTT GAGGGTGGTC AATGGAACTA TGGAGTAGGT 60

TGGACAGGAA CTTTTGGATA TTCTGATTAC TTACATTCTA CTCGATATCA TACAGCAACT 120

GTTAGACATG GGGGTAGAAC CTCTAAGGAT TATGCAAAAC CTGAGGCATG GGCTAGAGCT 180

TCCCTCACCA AGATTCCG 198

(2) INFORMATION FOR SEQ ID NO: 350:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 891 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 350:

GCTTCTTCTA TAGACAAAAA TATCATGGGT AAAATAATCA AGGCTATAGC TAGAAGGAGG 60

GACCAATCCA CTACTAATCC TAAGAACAAA ACACTCAAGA GAGCAGAAGA GAGAGGTTCA 120

CTGGCACTGA TAACGGCAAC CACCAAAGGA GAAACCAAGG ACACAGCCTT CATGGAAATG 180

AAAAAAGCAA AAGCCGTTCC AAAGAAAGCG ATAATGAGGC AAATCAAGAT ACTCCAAATA 240

TCAAGAGTAA AGGAAAGCTG ATAAACCGGC GAGAGGACAT TGCTAAACAA ACCTGCCAAA 300

ATCATCCCCC ACCCAACCGT AGGAACAAAA CCATAACGCT TAGCAAAAGG TTGGGGCAAG 360

ATAACATTAA ACATAACACC CATGGCACTC AGCAAACCTG TTATAAGAGC TAGCGGCGTC 420

ATGGATAACT GAGAGAGGTC TCCCTTTGTC GCCATCAAGC AAACACCCAG CATGGCAACC 480

AAAACATAGA AAACAGCGCT TTTTGACGCT CGTTTTTGAT AAACCAAGCG ATTGTAAAAG 540

AGGATAAAGA CAGGGCTAAT AAACTGTAAA ATAGTTGCTG TCGTAGCATT TGAGTATTCT 600

ACACAGAGAT AGAAAAAATA CTGAACTGAA AAAATCCCCA AAATAGCATA GGCTAAAAAG 660

GGCAGGTAAT TTTTCTTGTC TCGCCAAATA TCTAGCACTT GCGATTTTAA TTGTATTGCA 720

GACCAAATGA GTACAAGACT CCCTGCCAGT GTCAAACGCA TAGAGGTAAT CCAGCCCGAA 780

GACACCTGAT AATGAGTAAA GAAGTACTCT CCTAAAATTC CACAGATTCC CCATATTAAG 840

CCGGATAGGA GCGAATAAAT TTTTCCGTTA ACAATCTTTT TCTGATACTG A 891

(2) INFORMATION FOR SEQ ID NO: 351:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 325 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 351:

GAAAGCGTTC AATAGAACAT TGCTTTTTTA TTTTTAGAGT AAGCTAAGCG CTTCAGCATC 60

TGCGATGATG GTTACATCAG GGTGATTTTG GAGGCTACTT GCAGGTAGGT TCTCAGTCAC 120

TGGGCCAGAT ACTGTTCCGG CAATGGCTTC TGCTTTCGAC TCACCGTAAG CAAAAAGAAT 180

AATAGACTTG GCATCCAAAA TGTTTTTAAT CCCCATTGAA ATAGCTTGGG TTGGGACGTC 240

TTCAATCTTG GCAAAGAAGC GTGCATTGGC TTCGATAGTA GACTGGTCAA GTTCTACTAG 300

ATGCGTTTGA CTGTCAAATG GAGTG 325

(2) INFORMATION FOR SEQ ID NO: 352:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 344 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 352:

CAAGAGCAGT TTGATGATTT TTGATAAGCA TGCGAATTTA AAATACAAAT ATGGCAATCG 60

CAAGTTTTGG TGTAGAGGCT ATTATGTAGA TACGGTAGGC CGTAATCAGA AAGTGATAGC 120

TGAATATATT CAGAATCAAT TACAAGAAGA CAGAGTAGCA GACCTAGCTC ACGTTATTCG 180

AGTCAGTAGA TCCGTTTACT GGCGAAATAA ATAAGAGGAA GTAACGTnAA GTGCTTTAGC 240

ACCTGCTCGG GAAAGTGGTG CGCGAGGAAG CTATTTCAGG ATGCTTTGGC CCTGGCCGGT 300

AGAAGCGTTA TAGCCGCAGA CTACGACACT TCACACTGGT GGTT 344

(2) INFORMATION FOR SEQ ID NO: 353:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 692 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 353:

CCCTATCCCT GCTATTGGGG CTGCTCTCAT TGCTGCTTTG GCACAAATCA GTCTTCCAAT 60

TGGACCTGTT CCCTTCACTC TGCAAAACTT TGCAATCGGC TTGATTCTAC TGTCTTTAGA 120

CCGAGAGAGG CTGTACTTTC TGCTGGACTC TATCTTCTTC TAGGTGCTAT CGGTCTTCCT 180

GTCTTTGCAG GAGGTGGAGC TGGTTTTCAG GCTTTAGTTG GCCCTACTGC AGGCTATCTT 240

TGGTTTTATC TCGTTTACTC TGGACTTACT TCCTCTCTAA CCAACAGCAA GAGTGGTGTT 300

GTTAAGATTT TTCTTGCAAA CCTCTTGGGT GATGCCCTTG TCTTTGTCGG CGGGATTCTC 360

AGCTTGCATT TCCTAGCTGG AATGGCATTT GAAAAAGCTC TTGCTGTGGG GGTTCTTCCC 420

TTTATCATTC CAGACCTTGG CAAACTTCTA GCTATTAGTT TTATTAGCCG TCCCCTACTT 480

CAACGCCTTA AAAATCAGGC TTACTTTACT AACTAAAAAA GGATATCGAG TTATCATGAC 540

TCAATATCCT TTTCTTTTAT TTTGAAAACT TATACTCAAT GAAAATCAAA GAGCAAACTA 600

GGAAGCTAGC CGCAGGCTnG CAAAACACTG TTTTGAGGTT GTGGATGAAA CTGACGAGTA 660

AnATCTCATA CATACGGCAA GGCAAAGCTG AC 692

(2) INFORMATION FOR SEQ ID NO: 354:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1005 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 354:

GTGATGGACT ACTGGTTCAA AACGCATCCA GAAGATTTTT TCGATAATGT CGGACCTCTT 60

GTAGCCAGTA ACTTTTTTCA TACTTACACC GAAGATTTCC ACTTGATGAA GGAAATTGGA 120

GTTAATTCTT TCCGCACTTC CATCCAATGG AGTCGACTCA TCAAGAATTT AGAGACAGGT 180

GAGCCTGATC CAAAAGGTAT TGCTTTCTAC AATGCCATCA TTGAAGAAGC TAAAAAGAAC 240

CAGATGGATC TTGTGATGAA TTTACATCAT TTTGATTTAC CAGTGGAACT TCTTCAAAAA 300

TACGGTGGTT GGGAAAGCAA ACATGTAGTG GAGTTATTCG TGAAGTTTGC CAAGACTGCT 360

TTCACATGCT TTGGAGATAA GGTTCATTAC TGGACAACTT TCAATGAGCC AATGGTCATT 420

CCAGAAGCAG GGTACTTATA TGCTTTCCAT TATCCAAATC TAAAAGGAAA GGGAAAAGAG 480

GCCGTACAAG TCATCTATAA TCTAAACCTT GCTAGTGCAA AAGTGATTCA ACTATATCGC 540

TCATTAGAAC TTGATGGAAA GATTGGGATT ATTTTAAACT TGACACCTGC TTATCCAAGA 600

AGTAATTCTC CAGAAGACTT AGAAGCAAGT CGATTTACAG ATGACTTCTT TAACAAAGTC 660

TTCTTGAATC CAGCTGTTAA AGGAACTTTC CCAGAAAGAT TGGTAAAACA GCTAGAGAGA 720

GATGGCGTGT TATGGAGTCA TACCGAAAAA GAGCTTCAAC TGATGAAATC AAATACGGTT 780

GATTTTCTTG GAGTAAACTA CTACCATCCA AAACGTGTTC AAGCACAAGC AAATCCTGAG 840

GAATATCAGA CGCCCTGGAT GCCAGACCAA TACTTCAAAG AGTATGAATG GCTGGAGCGT 900

CGCATGAATC CATATCGTGG TTGGGAAATT TTTCCGAAAG CCATTTATGA TATTGCTATG 960

ATTGTGAAGG AAGAATATGG TAATATCCCA TGGTTTATCA GTGAA 1005

(2) INFORMATION FOR SEQ ID NO: 355:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 973 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 355:

CCGACAAGCA ATATTAAAAA GAGTAAACTA TTAACTAGTT AATTAACCGG TTTATTACTT 60

TATAGTGAAT CAAATATACT TAAGAAAAGA GGAAAGAATG AAAATTAATA AAAAATATCT 120

AGCAGGTTCA GTGGCAGTCC TTGCCCTAAG TGTTTGTTCC TATGAGCTTG GACGTTACCA 180

AGCTGGTCAG GATAAGAAAG AGTCTAATCG AGTTGCTTAT ATAGATGGTG ATCAGGCTGG 240

TCAAAAGGCA GAAAACTTGA CACCAGATGA AGTCAGTAAG AGGGAGGGGA TCAACGCCGA 300

ACAAATTGTT ATCAAGATTA CGGATCAAGG TTATGTGACC TCTCATGGAG ACCATTATCA 360

TTACTATAAT GGCAAGGTTC CTTATGATGC CATCATCAGT GAAGAGCTCC TCATGAAAGA 420

TCCGAATTAT CAGTTGAAGG ATTCAGACAT TGTCAATGAA ATCAAGGGTG GTTATGTCAT 480

TAAGGTAAAC GGTAAATACT ATGTTTACCT TAAGGATGCA GCTCATGCGG ATAATATTCG 540

GACAAAAGAA GAGATTAAAC GTCAGAAGCA GGAACGCAGT CATAATCATA ACTCAAGAGC 600

AGATAATGCT GTTGCTGCAG CCAGAGCCCA AGGACGTTAT ACAACGGATG ATGGGTATAT 660

CTTCAATGCA TCTGATATCA TTGAGGACAC GGGTGATGCT TATATCGTTC CTCACGGCGA 720

CCATTACCAT TACATTCCTA AGAATGAGTT ATCAGCTAGC GAGTTAGCTG CTGCAGAAGC 780

CTATTGGAAT GGGAAGCAGG GATCTCGTCC TTCTTCAAGT TCTAGTTATA ATGCAAATCC 840

AGCTCAACCA AGATTGTCAG AGAACCACAA TCTGACTGTC ACTCCAACTT ATCATCAAAA 900

TCAAGGGGGA AACATTTCAA GCCTTTTACG TGAATTGTAT GCTAACCCTT ATCAGAACGC 960

CATGTGGGAT CTG 973

(2) INFORMATION FOR SEQ ID NO: 356:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 843 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 356:

GGTCGCATCT GCAATATCTG TCGCCTCCAC ATAAGCGACA CCAGCCTTGT CTGCTGCCCG 60

TTTGACACGT TCTGCAGATT GACCCAGGAT GACCATCTTC TTGAGTCCAG TAATGTCTGG 120

CACCAATTCG TCAAACTCAT TGCCACGGTC CAAACCACCT GCAATCAAGA CGACCTTGCT 180

GTTGTCAAAT CCTGACAAGC TTTTTGAGTA GCCAAGATAT TAGTTGATTT ACTGTCGTTA 240

TAGAATTTAA CACsCTTGAT GTCATCCACA AACTGGAGAC GGTGTTTGAC ACCACCGAAG 300

GCTGAAAGAG TTTCCTTGAT GGTTTGATTG TCCACATCAC GAAGCTTGGC TACAGCAATA 360

GTCGCAAGGG CATTTTCCAC ATTGTGGCTA CCTGGAACAC CGATTTCATT CGCTGCCATG 420

ACTACTTCAC CACGGAAGTA GAGTTGACCA TCTTCCAGAT AAGCTCCATC AACCTTTTCA 480

AGTGTTGAAA ATGGTACAAC AGTGGCTTCT GTCTTGGAAG TCAAGTCTTT TGCCAAGTCT 540

TGATTAAAGT TCAAGACAAG GAAATCAGCT GCTGTCATCT TGTTCTGGAT ATTCCACTTG 600

GCTGCTACAT ATTCCGAAAA TGACCCATGG TAGTCGATAT GAGTTGGCAT GAGGTTGGTA 660

ATAACCGCAA TCTCTGGATG GAATTCTTGA ACACCCATGA GTTGGAAAGA AGAAAGTTCC 720

ATAACAAGCG TGTCCTTATC TGATGCTATT TGAGCAACCT GACTAGCTGG ATAGCCGATA 780

TTCCCTGATA AAAGACCATG TTGGCCAGCA GCAGTCAAAA CTTCCCGGGn TCCTCTAGAG 840

TCG 843

(2) INFORMATION FOR SEQ ID NO: 357:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 807 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 357:

TTTTTTTTAT ATTTTTTTTA TTTATTATTT TTTGGCAAAA AAGACCAATT TGCTTTGGAG 60

CATTGCTTCT GCATTAAATT GTCTATTTTT GCTCGTGCTG TTACGCTCTT TGTATCATGT 120

ATTAACTAGC AAGTGCAACT TGCAAACTAC TAGTAAGAGG AGAAAAACAA AATGGTTATG 180

ACTGACCCAA TCGCAGACTT CCTAACTCGT ATTCGTAATG CTAACCAAGC TAAACACGAA 240

GTACTTGAAG TACCTGCATC AAACATCAAA AAAGGGATTG CTGAAATCCT TAAACGCGAA 300

GGTTTTGTAA AAAACGTTGA AATCATTGAA GATGACAAAC AAGGCGTCAT CCGTGTATTT 360

CTTAAATACG GACCAAATGG TGAGAAAGTT ATCACTAACT TGAAACGTGT TTCTAAACCA 420

GGACTTCGTG TCTACAAAAA ACGTGAAGAC CTTCCAAAAG TTCTTAACGG ACTTGGAATT 480

GCCATCCTTT CAACTTCTGA AGGTTTGCTT ACTGATAAAG AAGCACGCCA AAAGAATGTT 540

GGTGGTGAGG TTATCGCTTA CGTTTGGTAA AATCAAGATA CAAAGCTCGT AAAGAACAAA 600

GCAAAATTAG GAAGTTGGAG AAGTTTGTTT ACAAACAGGC CAACTTATCT ATTTTGCACA 660

GTTCTTAGAG CGTGTTCAGT TCAGCTCTTG AGCTAAGTAA GTATCTGAAC CCCGTGAAAA 720

CTGGCCGTGC TGGCATGTTC GGGTAACAGG AGAnAATAAA CATGTCACGT ATTGGTAATA 780

AGTTCAGCTA AGGCCTTCGT AAAAGTT 807

(2) INFORMATION FOR SEQ ID NO: 358:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 653 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 358:

CCCAGTATTT TTGTCCAAGC ACGACCAGAA AAGGATGATA CAGATCTGGA ATTGGCTCTC 60

TTAACCATCT tTGAACAAAA TCCTCAGGCT CAGGTCACTA TTTTCGGTGC CTTGGGTGGC 120

CGTATTGACC ATATGTTGGC CAATGTCTTT CTGCCTAGCA ATCCTAAGTT GGCACCCTAT 180

ATGCATCAAA TAGAAATTGA GGATGGGCAA AACTTGATTA CTTATTGTCC AGAAGGAATC 240

AGTCAGCTAG AACCTCGTTC AGACTACGAC TATCTAGCCT TTATGCCAGT TCGGGATAGC 300

CAAGTATGAG TTGACAGAGG AAAATTTTTT CTTTAAAAAA GTGTACGCTT CTAACGAATA 360

TATAGATAGG GAAGTGTCGG TAACTTGCCC AGATGGTTAT GTGGTCGTAC TGCATAGCAA 420

GGACAGGAGG TAGGATGGAA AGTTTACTTA TTCTATTATT AATTGCCAAT CTAGCTGGTC 480

TCTTTCTGAT TTGGCAAAGG CAGGATAGGC AGGAGAAACA CTTAAGTAAG AGCTTGGAGG 540

ATCAGGCAGA TCATTTGTCA GACCAGCTGG ATTACCGCTT TGACCAAGCC AGACAAGCCA 600

GCCAGTTAGA CCAAAAAGAT TTGGAAGTGG TTGTCAGCGA CCGTTTGCAA GAA 653

(2) INFORMATION FOR SEQ ID NO: 359:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 641 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 359:

CACCATGTGA TGTGACGCTG GCCACAGCTG TCAGAAATCT GGCGAGCCAT CGTGTGCAAT 60

GACTCTTCCC GATGTAATCT TGTTCATAGT CCTTTGATGA ATATGTTCAA GCTGTAGAAG 120

GTGCGCTTCC TGAACACTTA TCAACTGTTA CAGGCGAGTT GACCAGTCAG GAAACAGATG 180

GCTGGTACAC ACTTGCCAAC ACTTCTTCAT CCCGCATTTA CCTAAAACAA GCCTTCCAAG 240

AAAATAGCAA CCTCCTAGAG CAAGTGGTAG AACCCTTGAC TATTATCACT GGTGGACACA 300

ACCACAAGGA CCAGTTGACC TATGCTTGGA AAACACTTTT GCAGAATGCG CCACATGATA 360

GTATCTGTGG CTGTAGCGTG GACGAAGTTC ACCGCGAGAT GGAAACGCGT TTTGCCAAGG 420

TCAACCAAGT AGGAAACTTT GTTAAAAGTA ACTTGCTCAA CGAGTGGAAG GGTAAAATTG 480

CTACGGATAA GGCTCAAAGT GACTATCTCT TTACTGTCAT TAACACAGGC TTGCATGATA 540

AGGTCGATAC TGTCAGCACA GTGATTGATG TGGCGACTTG TGATTTCAAG GAATTGCACC 600

CAACAGAAGG CTACAAAAAG ATGGCTGCTC TTATCTTGCC G 641

(2) INFORMATION FOR SEQ ID NO: 360:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1958 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 360:

CCTCAAGGCC AATTTGAAGG CTCTAAAACA ATGGAAAAGT GCTACACAGA TGTGACAGAA 60

TTTGCCATTC CAGCAAGTAC TCAAAAGCTT TACTTATCAC CAGTTTTAGA TGGCTTTAAT 120

AGCGAAATTA TTGCTTTTAA TCTTTCGACT TCACCCAACT TAGAACAAGT ACAAACAATG 180

TTAGAACAGG CATTCAAAGA GAAGCACTAC GAGAATACGA TTCTCCATAG TGACCAAGGC 240

TGGCAATATC AACACGATTC TTATCATCGG TTCCTAGAGA GTAAGGGAAT TCAAGCATCT 300

ATGTCACGCA AGGGCAACAG CCAAGACAAC GGTATGATGG AATCTTTCTT TGGCATTTTA 360

AAATCCGAAA TGTTTTATGG CTATGAGAAA ACATTTAAAT CACTTAACCA ATTGGAACAA 420

GCCATTATAG ACTATATTGA TTACTACAAC AACAAACGAA TTAAGGTAAA ACTAAAAGGA 480

CTTAGTCCTG TGCAGTACAG AACTAAATCC TTTGGATAAA TTAATTGTCT AACTTTTTGG 540

GGTCAGTACA AAACTCTTGC TACTATGCGT TTTATTATTG AAAGACTTAT TGGACTTTCT 600

CTCAAATCGA GTTTTTACTC AATTTTCTTA CTTGATTGGG ATTGAAATTC CAATTAATTT 660

CTCTGAGTAG AGTGTCTTGA TATTGGCTTC ATCAACAGAG GCCTTATCAA TTTTACGTTT 720

CAAGAAAAAT TCTTGAATGG TTTCGATTTC AGGCTCACGA ATAGCACGGT GTTTGTTTGA 780

GATGAGGATT TCATAGTGAA GCGGAGCTTG GGTAAAAATA ACATCTGTAT TCCCTGCAGA 840

ATAAACCTCA ACAAGGGTTG CATCGGTACT TTCTAGCTGA CTTTTTACAA GTTGCGAGTG 900

TGAGTTTGTC GTATTGATAA GCTTCATAAT ATTTCCTCCG ATTTTCTAAT TCTATTATAG 960

CACTTTTTGA ATAAAGTCGC TTGATTTATA CTCAATGAAA ATCAAAGAGC AAACTAGGAA 1020

GCTAGCCGCA GGCTATACTT GAGTACGGTA AGGCGACGCT GACGTGGTTT GAATTTTATT 1080

TTCGAAGAGT ATTAGCCAAT CTTATGCTGT TTTTTCCAAG ATTCAATGGC CCATTTATGG 1140

CTACCACGTT TAAGGTTTTT GATAGCCTCG TCAATAGGGA ACCAGGCAAT ATGATTAAAG 1200

TTTTCTAGTG GCTTTTGTAC TTCTTTGAAA GGAGTTGCTT CATAGAGGTA GGCAGGATTG 1260

TAGTAGTAGG TATCACGATG ACGAGAATAG AAATATTCGT CAGCTTGTCC GTAATAGGTA 1320

CCAATTTCTG CTGTGAAACC AAGCTCTTCA ATCAACTCAT GCTTTAGGGC TTCCTGATGA 1380

TTTTCACCTG CTTCAATTTC TCCACATGGT AGGAACCAAG CACCATTTGG TTCTTGAACA 1440

AGAACAATTT GTTTTTGTTC AGGATTAGGG ATAACTGCAT ATACGCCATA GCGAGCAATA 1500

TAGTCTGTAT TCACTTTTTT TCTCCGAAAG TTGGGTTTGC CATTGCATTT TCCTCATTAT 1560

CTAGTATCGT TATTATTATA GTGAAATGAA CCAAAAATAG TACACAATGT GGTATAATCT 1620

TCTTATGGCA TATTCAATAG ATTTTCGTAA AAAAGTTCTC TCTTATTGTG AGCGAACAGG 1680

TAGTATAACA GAAGCATCAC ACGTTTTCCA AATCTCACGT AATACCATTT ATGGCTGGTT 1740

AAAGCTAAAA GAGAAAACAG GAGAGCTAAA CCACCAAGTA AAAGGAATAA AACCAAGAAA 1800

GGTTGATAGA GATAGACTTA AAAACTATCT TACTGACAAT CCAGACGCTT ATTTGACTGA 1860

AATAGCTTCT GAATTTGGCT GTCATCCAAC TACCATCCAC TATGCGCTCA AAGCTATGGG 1920

tACACTCGAA AAAAAAAAGA ACTACACCTA CTATGAAC 1958

(2) INFORMATION FOR SEQ ID NO: 361:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 851 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 361:

TATGAAATTA AGTTATGATG ATAAAGTTCA GATCTATGAA CTTAGAAAAC AAGGATATAG 60

CTTAGAGAAG CTTTCAAATA AATTTGGGAT AmACAATTCT AATCTTAGGT ATATGATTAA 120

ATTGATTGAT CGTTACGGAA TAGAGTTCGT CAAAAAAGGA AAAAATCGTT ACTATTCTCC 180

TGATTTAAAA CAAGAAATGA TTAATAAAGT CTGACATGAA GGCTGGACTA AAGATAGAGT 240

TTCTCTTGAA TACGGTCTCC CAAGTCGTAC GATACTTCTT AACTGGCTAG CACAATACAG 300

GAAAAACGGG TATACTATTG TTGAGAAACC AAGAGGGAGA GTACCTGAGA GCGGAGAATG 360

CCATCCTAAA AAAGTTAAGA GAACTCCGAT TGAAGGAGGA AAAAGAGAAA GAAGAAAGAC 420

AGAAATTGTT TAAGAATTAA TGACTGAGTT TTCGTTAGAT CTTCTTTTAA AAGTCATTAA 480

ACTAGCTCGT TCGACCTACT ACTATCACTT GAAACAGCTA GATAAACCAG ATAAGGACCA 540

AGAGCTTAAA GCTGAAATTC AATCCATTTT TATCGAACAC AAAGGAAATT ATGCTTATCG 600

TCGGATTTAT TTAGAACTAA GAAATCGTGG TTATCTGGTA AATCATAAAA GAGTTCAAGG 660

CTTGATGAAA GTACTCAATt TACAAGCTAA AACGCGACAG AAACGAAAAT ATTCTTCTCA 720

TAAAGGAGAC GTTGGCAAGA AGGCAGAGAA TCTCATTCAA GGCCAATTTG AAGGCTCTAA 780

AACAATGGAA CAGTGCTACA CAGATGTGAC AGAATTTGCC ATTCCAGTAA GTACTTAAAA 840

GCTTTACTTA T 851

(2) INFORMATION FOR SEQ ID NO: 362:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1168 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 362:

GGGTAGAATC GATATCTCCA ATGAGTTGGT tTAGCTGGTG AAACTGTAAA AAGATTTCGw 60

CCAATTCAAG GTTGAGGCAT CGCAAACTAT GGACTGTTTC CTCGTCAGTT CTGGAAAGAA 120

AACGGGATAA GGTTGGCTGT GAAGCAAGCT GCCCTCCTTC CAACAATTTT GGAAAGTAGG 180

CATCAGCTGA CAATTCTTTA CAAGCATAGT CCGTTCCATA ACCTGTTAAC AGTTGAAAGA 240

GGAACTGGAC AAGGATATCT GAATCCGAAT AACGACAGTA GCGGCGTTGG TCATTCGTTA 300

CTAAATACTT AGAAATCCGC TCTTTTAGTT TCAACTGGGA AAAAAGTTCC TGAAAAAAGA 360

TAAGACCACC ATACTGGGTT AAATGACCTC CATCGAAAGA TAGTTGGTAA AAAGACTTGT 420

TTTGGAAGTG ATGATTTGGT AAACTGTTCA TGTGAGTTTC CTTTCTTTTT GTGTTTTTTT 480

CTACACTTAT ACCATAAAGG GGAAACTCTT TTTTGTCTAG TAAAAAACAC CCATTGGGTG 540

AAAAAAGAAA CCATCCAGGA TCTAAGCTAA GGCAAGGATT CTGGATGGTT TTTAGATTTG 600

GGGTGAATAA TTGGGGTTTT AGCTGCTTGC GGCCAATCAG GTTCAGATAC AAAAACTTAC 660

TCATCAACCT TTAGTGGAAA TCCAACTACA TTTAACTATC TATTAGACTA TTACGCTGAT 720

AATATAGTCA ATTGAAACAA GAACAAGACA AAAGAGCCTC ATAAAAGGTA TTGCAACTTG 780

GTAATACCTT TTTGAGGTGC TTTTTGATAT GAGCCCATGT TTTCTCAATA GGATTGTACT 840

CAGGTGAGTA GGGAGGAAGA GGTAAAAGTT TATACCCAAA CTCTTCACAC AAGAGTTCTA 900

ACTTACCCAT TCTATGGAAT CTTGCATTAT CCATAATAAT AACCGATGGT GTGTTTAATG 960

TTGGTAAGAG AAATTTCTGA AACCAAGCTT CAAAAAAGTC GCTCGTCATC GTCTCTTCGT 1020

AAGTTATTGG AGCGATTAAC TCACCATTTG TTAGACCTGC AACCAAAGAA ATCCTCTGAT 1080

ATCTTCTTCC AGATACTTTG CCTCTTCTTA ACTGACCTTT TAATGAGCGA CCATATTCTC 1140

GATAAAAATA AGTATCGAAT CCTGTTTC 1168

(2) INFORMATION FOR SEQ ID NO: 363:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4483 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 363:

GTCAGCTTCA GCAAGCCCAT CAGCTTCTGA ATCTGCATCA ACCAGTGCGT CCGCTTCAGC 60

GTCAACCAGT GCGTCGGCTT CAGCGTCGAC AAGTGCTTCG GCTTCAGCAT CAACGAGTGC 120

GTCGGCCTCA GCAAGCGCAA GTACCTCAGC GTCAGCTTCC GCCTCAACCA GTGCGTCGGC 180

TTCAGCAAGC ACAAGTGCGT CAGCCTCAGC AAGTATCTCA GCGTCTGAAT CGGCATCAAC 240

GAGTGCGTCT GAGTCAGCAT CAACGAGTAC GTCAGCCTCA GCAAGCACAT CAGCTTCTGA 300

ATCTGCATCA ACCAGTGCGT CAGCCTCAGC ATCGACAAGC GCCTCAGCTT CAGCAAGTAC 360

CAGTGCTTCA GCCTCAGCGT CGACAAGTGC GTCGGCCTCA ACCAGTGCAT CTGAATCGGC 420

ATCAACCAGT GCGTCAGCCT CAGCAAGTAC TAGTGCATCA GCTTCAGCAT CAACGAGTGC 480

ATCGGCTTCA GCATCAACCA GTGCCTCGGC TTCAGCGTCA ACCAGTGCGT CAGCTTCAGC 540

AAGTACCAGT GCTTCAGTCT CAGCATCAAC AAGTGCTTCA GCCTCAGCAT CGACAAGTGC 600

CTCGGCTTCA GCAAGCACAT CAGCATCTGA ATCAGCGTCG ACAAGCGCCT CAGCTTCAGC 660

AAGTACCAGT GCGTCAGCCT CAGCGTCGAC AAGTGCGTCA GCCTCAGCAA GTACTAGTGC 720

ATCAGCTTCA GCATCAACGA GTGCATCGGC TTCGGCGTCA ACCAGTGCAT CAGAGTCAGC 780

AAGTACCAGT GCGTCAGCTT CCGCATCAAC AAGTGCCTCG GCTTCAGCAA GCACCAGTGC 840

GTCGGCTTCA GCAAGTACTA GCGCCTCAGC CTCAGCCTCA ACCAGTGCGT CAGCCTCAGC 900

AAGTATCTCA GCGTCTGAAT CGGCATCAAC GAGTGCGTCC GCTTCAGCAA GTACTAGCGC 960

CTCAGCCTCA GCGTCAACAA GTGCATCGGC TTCAGCGTCA ACGAGTGCGT CTGAATCGGC 1020

ATCAACGAGT GCGTCCGCTT CAGCAAGTAC TAGCGCCTCA GCCTCAGCGT CAACAAGTGC 1080

ATCGGCTTCA GCATCAACGA GTGCGTCCGC TTCAGCAAGT ACTAGCGCCT CAGCCTCAGC 1140

GTCAACAAGT GCATCGGCTT CAGCGTCAAC GAGTGCGTCT GAGTCAGCAT CAACGAGTGC 1200

GTCAGCCTCA GCAAGCACAT CAGCTTCTGA ATCTGCATCA ACCAGTGCGT CAgCCTCAGC 1260

ATCGACAAGC GCCTCAGCTT CAGCAAGTAC CAGTGCGTCA GcTCAGCGTC GACAAGTGCs 1320

TCrGCTTCAG CAAGTACCAG TGCGTCAGCC TCAGCAAGTA CCAGTGCkTC AGCCTCAGCG 1380

TCGACAAGTG CGTCGGCCTC AACCAGTGCA TCTGAATCGG CATCAACCAG TGCGTCAGCC 1440

TCAGCAAGTA CTAGCGCCTC AGCCTCAGCA TCAACGAGTG CGTCCGCTTC AGCAAGTACT 1500

AGTGCATCAG CTTCAGCAAG TACTAGCGCC TCAGCCTCAG CGTCGACAAG CGCCTCAGCT 1560

TCAGCAAGTA CCAGTGCGTC AGCCTCAGCG TCGACAAGTG CGTCGGCTTC AGCAAGTACC 1620

TCAGCGTCTG AATCAGCATC AACAAGTGCG TCGGCTTCAG CATCAACGAG TGCATCAGCT 1680

TCAGCATCAA CAAGTGCTTC AGCTTCAGCA AGTACCAGTG CGTCGGCTTC AGCATCAACG 1740

AGTGCTTCAG TCTCAGCGTC AACCAGTGCC TCTGAATCCG CATCAACAAG TGCCTCGGCT 1800

TCAGCAAGCA CCAGTGCTTC GGCTTCAGCG TCAACGAGTG CGTCTGAGTC AGCATCAACG 1860

AGTGCGTCAC CTCAGCAAGC ACATCAGCTT CTGAATCTGC ATCAACCAGT GCGTCACTTC 1920

CGCATCAACA AGCGCCTCGG CCTCAGCAAG TACAAGTGCT TCAGCCTCAG CATCAACCAG 1980

TGCATCAGCT TCAGCCTCAA CAAGTGCTTC AGCCTCAGCG TCAACCAGTG CCTCGGCTTC 2040

AGCAAGTACC AGTGCGTCAG cTTCAGCAAG CACAAGTGCG TCAGCTTCAG CATCAACCAG 2100

TGCTTCGGCT TCGGCATCAA CAAGTGCCTC AGCATCAGCA TCAACGAGTG CGTCAsCTCA 2160

GCAAGTACTA GTGCATCAGC ATCAGCATCA ACCAGTGCAT CAGCCTCAGC AAGTATCTCA 2220

GCGTCTGAAT CGGCATCAAC GAGTGCATCA GCATCAGCAT CAACGAGTGC ATCGGCTTCA 2280

GCGTCAACCA GTGCATCAGT CTCAGCAAGC ACCAGTGCGT CGGCTTCAGC ATCAACCAGT 2340

GCCTCAGCCT CAGCAAGTAT CTCAGCGTCT GAATCGGCAT CAACGAGTGC GTCAGcCTCA 2400

GCAAGTACTA GTGCATCAGC ATCAGCATCA ACGAGTGCAT CGGCTTCAGC AAGTACCAGC 2460

GCCTCAGCTT CAGCAAGCAC CAGTGCGTCA GCCTCAGCAA GTACCAGCGC CTCAGCCTCA 2520

GCAAGCACCA GTGCCTCAGC TTCAGCAAGT ACCAGTGCGT CAGCCTCAGC GTCGACAAGT 2580

GCGTCGGCTT CAGCAAGTAC CTCAGCGTCT GAATCAGCAT CAACGAGTGC ATCAGCTTCA 2640

GCATCAACAA GTGCTTCAGC TTCAGCAAGT ACCAGTGCGT CGGCTTCAGC ATCAACGAGT 2700

GCTTCAGTCT CAGCGTCAAC CAGTGCCTCT GAATCAGCAT CAACAAGTGC CTCGGCTTCA 2760

GCAAGCACCA GTGCGTCGGC TTCAGCAAGT ACTAGTGCAT CGGCTTCAGC ATCGACAAGT 2820

GCGTCTGAAT CGGCATCAAC GAGTGCTTCG GCTTCAGCAT CAACGAGTGC GTCAGCCTCA 2880

GCAAGCACAT CAGCTTCTGA ATCTGCATCA ACCAGTGCGT CCGCTTCAGC GTCAACCAGT 2940

GCGTCGGCTT CAGCGTCGAC AAGTGCTTCG GCTTCAGCAT CAACGAGTGC GTCGGCCTCA 3000

GCAAGCGCAA GTACCTCAGC GTCAGCTTCC GCCTCAACCA GTGCGTCCGC TTCAGCAAGC 3060

ACAAGTGCGT CAGCCTCAGC AAGTATCTCA GCGTCTGAAT CGGCATCAAC GAGTGCGTCG 3120

GCCTCAGCAA GCGCAAGTAC CTCAGCGTCA GCTTCCGCCT CAACCAGTGC GTCGGCTTCA 3180

GCAAGCACAA GTGCGTCAGC CTCAGCAAGT ATCTCAGCGT CTGAATCGGC ATCAACGAGT 3240

GCGTCTGAGT CAGCATCAAC GAGTACGTCA GCCTCAGCAA GCACATCAGC TTCTGAATCG 3300

GCATCAACCA GTGCGTCAGC CTCAGCATCG ACAAGCGCCT CAGCTTCAGC AAGTACCAGT 3360

GCTTCAGCCT CAGCGTCGAC AAGTGCGTCG GCCTCAACCA GTGCATCTGA ATCGGCATCA 3420

ACCAGTGCGT CAGCCTCAGC AAGTACTAGT GCATCAGCTT CAGCATCAAC GAGTGCATCG 3480

GCTTCAGCAT CAACCAGTGC CTCGGCTTCA GCGTCAACCA GTGCGTCAGC TTCAGCAAGT 3540

ACCAGTGCTT CAGTCTCAGC ATCAACAAGT GCTTCAGCCT CAGCATCGAC AAGTGCCTCG 3600

GCTTCAGCAA GCACATCAGC ATCTGAATCA GCGTCGACAA GCGCCTCAGC TTCAGCAAGT 3660

ACCAGTGCGT CAGCCTCAGC GTCGACAAGT GCGTCAGCCT CAGCAAGTAC TAGTGCATCA 3720

GCTTCAGCAT CAACGAGTGC ATCGGCTTCG GCGTCAACCA GTGCATCAGA GTCAGCAAGT 3780

ACCAGTGCGT CAGCTTCCGC ATCAACAAGT GCCTCGGCTT CAGCAAGCAC CAGTGCGTCG 3840

GCTTCAGCAA GTACTAGCGC CTCAGCCTCA GCCTCAACCA GTGCGTCAGC CTCAGCAAGT 3900

ATCTCAGCGT CTGAATCGGC ATCAACGAGT GCGTCCGCTT CAGCAAGTAC TAGCGCCTCA 3960

GCCTCAGCGT CAACAAGTGC ATCGGCTTCA GCGTCAACGA GTGCGTCTGA ATCGGCATCA 4020

ACGAGTGCGT CCGCTTCAGC AAGTACTAGC GCCTCAGCCT CAGCGTCAAC AAGTGCATCG 4080

GCTTCAGCAT CAACGAGTGC GTCCGCTTCA GCAAGTACTA GCGCCTCAGC CTCAGCGTCA 4140

ACAAGTGCAT CGGGTTCAGC GTCAACGAGT GCGTCTGAGT CAGCATCAAC GAGTGCGTCA 4200

CCTCAkCAAG CACATCAGCT TCTGAATCTG CATCAACCAG TGCGTCACTT CCGCATCAAC 4260

AAGCGCCTCG GCCTCAGCAA GTACAAGTGC TTCAGCCTCA GCATCAACCA GTGCATCAGC 4320

TTCAGCCTCA ACAAGTGCTT CAGCCTCAGC GTCAGACCAG TGCCTCGGCT TCAGCAAGTA 4380

CCAGTGCGTC ACTTCAGCAA GCACAAGTGC GTCAGCTTCA GCATCAACCA GTGCTTCGGC 4440

TTCGGCATCA ACAAGTGCCT CAGCATCAGC ATCAACGAGT GCG 4483

(2) INFORMATION FOR SEQ ID NO: 364:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2550 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 364:

GTACCTCAGC GTCCTTCCGC CTCAACCAGT GCGTCCGCTT CAGCAAGCAC AAGTGCGTCA 60

CCTCAGCAAG TATCTCAGCG TCTGAATCGG CATCAACGAG TGCGTCGGCC TCAGCAAGCG 120

CAAGTACCTC AGCGTCACTT CCGCCTCAAC CAGTGCGTCG GCTTCAGCAA GCACAAGTGC 180

GTCAsCTCAG CAAGTATCTC AGCGTCTGAA TCGGCATCAA CGAGTGCGTC TGAGTCAGCA 240

TCAACGAGTA CGTCAGCCTC AGCAAGCACA TCAGCTTCTG AATCGGCATC AACCAGTGCG 300

TCAGCCTCAG CATCGACAAG CGCCTCAGCT TCAGCAAGTA CCAGTGCTTC AGCCTCAGCG 360

TCGACAAGTG CGTCGGCCTC AACCAGTGCA TCTGAATCGG CATCAACCAG TGCGTCAGCC 420

TCAGCAAGTA CTAGTGCATC AGCTTCAGCA TCAACGAGTG CATCGGCTTC AGCATCAACC 480

AGTGCCTCGG CTTCAGCGTC AACCAGTGCG TCAGCTTCAG CAAGTACCAG TGCTTCAGTC 540

TCAGCATCAA CAAGTGCTTC AGCCTCAGCA TCGACAAGTG CCTCGGCTTC AGCAAGCACA 600

TCAGCATCTG AATCAGCGTC GACAAGTGCG TCGGCCTCAA CCAGTGCATC TGAATCGGCA 660

TCAACCAGTG CGTCAGCCTC AGCAAGTACT AGTGCATCAG CTTCAGCATC AACGAGTGCA 720

TCGGCTTCGG CGTCAACCAG TGCATCAGAG TCAGCAAGTA CCAGTGCGTC AGCTTCCGCA 780

TCAACAAGTG CCTCGGCTTC AGCAAGCACA TCAGCATCTG AATCAGCGTC AACCAGTGCT 840

TCGGCTTCAG CAAGTACCAG TGCTTCAGCT TCAGCATCAA CCAGCGCCTC GGCCTCAGCA 900

AGCACCTCAG CTTCTGAATC GGCCTCAACC AGCGCCTCGG CCTCAGCAAG CACCTCAGCT 960

TCTGAATCGG CCTCAACCAG CGCCTCAGCC TCAGCATCAA CGAGTGCTTC GGCTTCAGCA 1020

AGCACAAGCG CCTCGGGTTC AGCATCAACG AGTACGTCAG CTTCAGCGTC AACCAGTGCT 1080

TCAGCCTCAG CATCAACAAG TGCGTCAGCC TCAGCAAGTA TCTCAGCGTC TGAATCGGCA 1140

TCAACGAGTG CGTCTGAGTC AGCATCAACG AGTACGTCAG CCTCAGCAAG CACAAGTGCT 1200

TCAGCCTCAG CAAGTATCTC AGCGTCTGAA TCGGCATCAA CGAGTGCGTC CGCTTCAGCA 1260

AGTACTAGCG CCTCAGCATC AGCGTCAACA AGTGCTTCGG CTTCAGCGTC AACGAGTGCG 1320

TCTGAGTCAG CATCAACGAG TACGTCAGCC TCAGCAAGCA CATCAGCTTC TGAATCTGCA 1380

TCAACCAGTG CGTCAGCCTC AGCATCGACA AGCGCCTCAG CTTCAGCAAG TACCAGTGCG 1440

TCAGCCTCAG CAAGTACCAG TGCTTCAGCC TCAGCGTCGA CAAGTGCGTC GGCCTCAACC 1500

AGTGCATCTG AATCGGCATC AACCAGTGCG TCAGCTCAGC AAGTACTAGT GCATCAGCTT 1560

CAGCATCAAC GAGTGCATCG GCTTCGGCGT CAACCAGTGC ATCAGAGTCA GCAAGTACCA 1620

GTGCGTCACt TCCGCATCAA CAAGTGCCTC GGCTTCAGCA AGCACATCAG CATCTGAATC 1680

AGCGTCAACC AGTGCTTCGG CTTCAGCAAG TACCAGTGCT TCAGCTTCAG CATCAACCAG 1740

1380

CGCCTCGGCC TCAGCAAGCA CCTCAGCTTC TGAATCGGCC TCAACCAGCG CCTCGGCCTC 1800

AGCAAGCACC TCAGCTTCTG AATCGGCCTC AACCAGCGCC TCAGCCTCAG CATCAACGAG 1860

TGCTTCGGCT TCAGCAAGCA CAAGCGCCTC GGGTTCAGCA TCAACGAGTA CGTCAGCTTC 1920

AGCGTCAACC AGTGCTTCAG CCTCAGCATC AACAAGTGCG TCAGCCTCAG CAAGTATCTC 1980

AGCGTCTGAA TCGGCATCAA CGAGTGCGTC TGAGTCAGCA TCAACGAGTA CGTCAGCCTC 2040

AGCAAGCACC TCAGCTTCTG AATCGGCCTC AACCAGTGCG TCAGCCTCAG CATCGACAAG 2100

CGCCTCAGCT TCAGCAAGTA CCAGTGCTTC AGCCTCAGCG TCGACAAGTG CGTCGGCCTC 2160

AACCAGTGCA TCTGAATCGG CATCAACCAG TGCGTCAGCC TCAGCAAGTA CTAGTGCATC 2220

GGCTTCAGCA TCAACCAGTG CCTCGGCTTC AGCGTCAACC AGTGCGTCAG CTTCAGCAAG 2280

TACCAGTGCT TCAGTCTCAG CATCAACAAG TGCTTCAGCC TCAGCATCGA CAAGTGCCTC 2340

GGCTTCAGCA AGCACATCAG CATCTGAATC AGCGTCGACA AGCGCCTCAG CTTCAGCAAG 2400

TACCAGTGCG TCAGCCTCAG CGTCGACAAG TGCGTCAGCT ACAGCAAGTA CTAGTGCATC 2460

AGCTTCAGCA TCAACGAGTG CATCGGCTTC GGCGTCAACC AGTGCATCAG AGTCAGCAAG 2520

TACCAGTGCG TCAGTTCACG CATCAACAAG 2550

(2) INFORMATION FOR SEQ ID NO: 365:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1436 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 365:

ACCCAGCAAG TACTAGTGCA TCGGCTTCAG CAAGCACCAG TGCGTCGGCT TCAGCATCAA 60

CCAGTGCCTC AGCCTCAGCA AGTATCTCAG CGTCTGAATC GGCATCAACG AGTGCGTCAC 120

CTCAGCAAGT ACTAGTGCAT CAGCATCAGC ATCAACGAGT GCATCGGCTT CAGCAAGTAC 180

CAGCGCCTCA GCTTCAGCAA GCACCAGTGC GTCAsCTCAG CAAGTACCAG CGCCTCAGCC 240

TCAGCAAGCA CCAGTGCCTC AGCTTCAGCA AGTACCAGTG CGTCAGCCTC AGCGTCGACA 300

AGTGCGTCGG CTTCAGCAAG TACCTCAGCG TCTGAATCAG CATCAACGAG TGCATCAGCT 360

TCAGCATCAA CAAGTGCTTC AGCTTCAGCA AGTATCTCAG CGTCTGAATC GGCATCAACG 420

AGTGCGTCCG CTTCAGCAAG TACTAGCGCC TCAGCATCAG CGTCAACAAG TGCTTCGGCT 480

TCAGCGTCAA CGAGTGCGTC TGAGTCAGCA TCAACGAGTA CGTCAGCCTC AGCAAGCACA 540

TCAGCTTCTG AATCTGCATC AACCAGTGCG TCAGCCTCAG CATCGACAAG CGCCTCAGCT 600

TCAGCAAGTA CCAGTGCGTC AgCCTCAGCA AGTACCAGTG CTTCAGCCTC AGCGTCGACA 660

AGTGCGTCGG CCTCAACCAG TGCATCTGAA TCGGCATCAA CCAGTGCGTC AGCCTCAGCA 720

AGTACTAGCG CCTCAGCCTC AGCATCAACG AGTGCGTCCG CTTCAGCAAG TACTAGTGCA 780

TCAGCTTCAG CAAGTACTAG CGCCTCAGCC TCAGCGTCGA CAAGCGCCTC AGCTTCAGCA 840

AGTACCAGTG CGTCAGCCTC AGCGTCGACA AGTGCGTCGG CTTCAGCAAG TACCTCAGCG 900

TCTGAATCAG CATCAACAAG TGCGTCGGCT TCAGCATCAA CGAGTGCATC AGCTTCAGCA 960

TCAACAAGTG CTTCAGCTTC AGCAAGTACC AGTGCGTCGG CTTCAGCATC AACGAGTGCT 1020

TCAGTCTCAG CGTCAACCAG TGCCTCTGAA TCCGCATCAA CAAGTGCCTC GGCTTCAGCA 1080

AGCACCAGTG CTTCGGCTTC AGCGTCAACG AGTGCGTCTG AGTCAGCATC AACGAGTGCG 1140

TCAGCCTCAG CAAGCACATC AGCTTCTGAA TCTGCATCAA CCAGTGCGTC AGCTTCCGCA 1200

TCAACAAGCG CCTCGGCCTC AGCAAGTACA AGTGCTTCAG CCTCAGCATC AACCAGTGCA 1260

TCAGCTTCAG CCTCAACAAG TGCTTCAGCC TCAGCGTCAA CCAGTGCCTC GGCTTCAGCA 1320

AGTACCAGTG CGTCAGCTTC AGCAAGCACA AGTGCGTCAG CTTCAGCATC AACCAGTGCT 1380

TCGGCTTCGG CATCAACAAG TGCCTCAGCA TCAGCATCAA CGAGTGCGTC AGCCGG 1436

(2) INFORMATION FOR SEQ ID NO: 366:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 735 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 366:

GCAGTTGCCA CACCGTGCTG ACCAGCACCC GTTCCTGCGA TAATTTTCTT TTTACCCATG 60

CGTwTGGCAA GCCAAACTTG TCCTAAGGCA TTGTTAATCT TGTGGGCTCC TGTATGGTTA 120

AGGTCTTCCC GTTTGAGATA AATCTTGCTC CGCCAATATG CTGGGTCAAG TTTTTTGCGT 180

AATAAAGAGG AGTTTCACGT CCTACGTACT GGCGCAAAAG CTGGTTTAAT TCCTCTTGGA 240

AACTTGGGTC TGCCTGACTT TCACGGTAGG CCTTCTCCAA CTCCAAAACT GCTGTCATCA 300

ATGTTTCTGG GACAAAACGT CCGCCGAATT TTCCGTAAAA TCCATCTTTA TTTGGTTCCT 360

GATATGCCAT GCTTTACCCT CTCTATAAAT CTTCTAATCT TTTCATGATC TTTTTGTCCA 420

TCTGTCTCCA CTCCGCTCGA TACATCTACT GCATAGGGAG TAAAGTGTTG AATTGCTTTT 480

ACTACATTAT CTTCATTAAG GCCACCTGCG ATAAAGAAGG GCTGTGCTAG TCCAGTCGTA 540

TCCAGTTGAC CCCAATCAAA GGGCTGGCCA CTTCCTGCCA CAGGGGCATC AAAGAGTAGA 600

TAATCTGCCT GAGAATTGGG GACATGCCCA TTTCCATCTA CCTGCACAGC CTGAATACTG 660

GCACAAGGCA AATTCTCAAA TAAATCATCT GCCACCTGAC CGTGAACTTG AACCAAGTCC 720

AAGCCGGGGA TCCTC 735

(2) INFORMATION FOR SEQ ID NO: 367:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1702 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 367:

TACTAGCGCC TCAGCCTCAG CGTCAACAAG TGCATCGGCT TCAGCATCAA CGAGTGCGTC 60

CGCTTCAGCA AGTACTAGCG CCTCAGCCTC AGCGTCAACA AGTGCATCGG CTTCAGCGTC 120

AACGAGTGCG TCTGAGTCAG CATCAACGAG TGCGTCAGCC TCAGCAAGCA CATCAGCTTC 180

TGAATCTGCA TCAACCAGTG CGTCAGCCTC AGCATCGACA AGCGCCTCAG CTTCAGCAAG 240

TACCAGTGCG TCAGCCTCAG CGTCGACAAG TGCGTCGGCT TCAGCAAGTA CCAGTGCGTC 300

AGCCTCAGCA AGTACCAGTG CGTCAGCCTC AGCGTCGACA AGTGCGTCGG CCTCAACCAG 360

TGCATCTGAA TCGGCATCAA CCAGTGCGTC AGCCTCAGCA AGTACTAGTG CATCAGCTTC 420

AGCATCAACG AGTGCATCGG CTTCAGCATC AACCAGTGCA TCAGAGTCAG CAAGTACCAG 480

TGCGTCAgCT TCCGCATCAA CAAGTGCCTC GGCTTCAGCA AGTACTAGCG CCTCAGCCTC 540

AGCGTCAACA AGTGCTTCAG CTTCCGCGTC AACCAGCGCC TCGGCCTCAG CAAGTATCTC 600

AGCGTCTGAA TCGGCATCAA CAAGTGCCTC GGCTTCAGCA TCAACGAGTG CATCAGTCTC 660

AGCAAGCACC AGTGCGTCGG CCTCAGCAAG CACCAGCGCG TCTGAATCCG CATCAACCAG 720

TGCCTCAGCT TCAGCAAGTA CCTCAGCATC TGAATCAGCA TCAACAAGTG CATCGGCTTC 780

AGCAAGCACA AGTGCTTCAG CCTCAGCAAG TATCTCAGCG TCTGAATCGG CATCAACGAG 840

TGCGTCCGCT TCAGCAAGTA CTAGCGCCTC AGCATCAGCG TCAACAAGTG CTTCGGCTTC 900

AGCGTCAACG AGTGCGTCTG AGTCAGCATC AACGAGTACG TCAGCCTCAG CAAGCACATC 960

AGCTTCTGAA TCTGCATCAA CCAGTGCGTC AGCCTCAGCA TCGACAAGCG CCTCAGCTTC 1020

AGCAAGTACC AGTGCGTCAG CCTCAGCAAG TACCAGTGCT TCAGCCTCAG CGTCGACAAG 1080

TGCGTCGGCC TCAACCAGTG CATCTGAATC GGCATCAACC AGTGCGTCAG CCTCAGCAAG 1140

TACTAGCGCC TCAGCCTCAG CATCAACGAG TGCGTCCGCT TCAGCAAGTA CTAGTGCATC 1200

AGCATCAGCA TCAACGAGTG CATCGGCTTC AGCAAGTACC AGCGCCTCAG CTTCAGCAAG 1260

CACCAGTGCG TCAGCCTCAG CAAGTACCAG CGCCTCAGCC TCAGCAAGCA CCAGTGCCTC 1320

AGCTTCAGCA AGTACCAGTG CGTCAGCCTC AGCGTCGACA AGTGCGTCGG CTTCAGCAAG 1380

TACCTCAGCG TCTGAATCAG CATCAACGAG TGCATCAGCT TCAGCATCAA CAAGTGCTTC 1440

AGCTTCAGCA AGTACCAGTG CGTCGGCTTC AGCATCAACG AGTGCTTCAG TCTCAGCGTC 1500

AACCAGTGCC TCTGAATCAG CATCAACAAG TGCCTCGGCT TCAGCAAGCA CCAGTGCGTC 1560

GGCTTCAGCA AGTACTAGTG CATCGGCTTC AGCATCGACA AGTGCGTCTG AATCGGCATC 1620

AACGAGTGCT TCGGCTTCAG CATCAACGAG TGCGTCAGCC TCAGCAAGCA CATCAGCTTC 1680

TGAATCTGCA TCAACCAGTG CG 1702

(2) INFORMATION FOR SEQ ID NO: 368:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 941 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 368:

ACCAGTGCAT CAGCTTCAGC CTCAACAAGT GCTTCAGCCT CAGCGTCAAC CAGTGCCTCG 60

GCTTCAGCAA GTACCAGTGC GTCACTTCAG CAAGCACAAG TGCGTCACTT CAGCATCAAC 120

CAGTGCTTCG GCTTCGGCAT CAACAAGTGC CTCAGCATCA GCATCAACGA GTGCGTCACC 180

TCAGCAAGTA CTAGTGCATC AGCATCAGCA TCAACCAGTG CATCAGCCTC AGCAAGTATC 240

TCAGCGTCTG AATCGGCATC AACGAGTGCA TCAGCATCAG CATCAACGAG TGCATCGGCT 300

TCAGCGTCAA CCAGTGCATC AGTCTCAGCA AGCACCAGTG CGTCGGCTTC AGCATCAACG 360

AGTGCCTCAG CCTCAGCAAG TATCTCAGCG TCTGAATCGG CATCAACGAG TGCGTCAGCC 420

TCAGCAAGTA CTAGTGCATC GGCTTCAGCA AGCACCAGTG CGTCGGCTTC AGCATCAACC 480

AGTGCCTCAG CCTCAGCAAG TATCTCAGCG TCTGAATCGG CATCAACGAG TGCGTCAGCC 540

TCAGCAAGTA CTAGTGCATC AGCATCAGCA TCAACGAGTG CATCGGCTTC AGCAAGTACC 600

AGCGCCTCAG CTTCAGCAAG CACCAGTGCG TCAGCCTCAG CAAGTACCAG CGCCTCAGCC 660

TCAGCAAGCA CCAGTGCCTC AGCTTCAGCA AGTACCAGTG CGTCAGCCTC AGCGTCGACA 720

AGTGCGTCGG CTTCAGCAAG TACCTCAGCG TCTGAATCAG CATCAACGAG TGCATCAGCT 780

TCAGCATCAA CAAGTGCTTC AGCTTCAGCA AGTACCAGTG CGTCGGCTTC AGCATCAACG 840

AGTGCTTCAG TCTCAGCGTC AACCAGTGCC TCTGAATCAG CATCAACAAG TGCCTCGGCT 900

TCAGCAAGCA CCAGTGCGTC GGCTTCAGCA AGTACTAGTG C 941

(2) INFORMATION FOR SEQ ID NO: 369:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 869 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 369:

CAGCAAGTAC TAGTGCATCA GCTTCAGCAT CAACGAGTGC ATCGGCTTCT GCGTCAACCA 60

GTGCATCAGA GTCAGCAAGT ACCAGTGCGT CAGCTTCCGC ATCAACAAGT GCCTCGGCTT 120

CAGCAAGCAC CAGTGCGTCG GCTTCAGCAA GTACTAGCGC CTCAGCCTCA GCCTCAACCA 180

GTGCGTCAGC CTCAGCAAGT ATCTCAGCGT CTGAATCGGC ATCAACGAGT GCGTCCGCTT 240

CAGCAAGTAC TAGCGCCTCA GCCTCAGCGT CAACAAGTGC ATCGGCTTCA GCGTCAACGA 300

GTGCGTCTGA ATCGGCATCA ACGAGTGCGT CCGCTTCAGC AAGTACTAGC GCCTCAGCCT 360

CAGCGTCAAC AAGTGCATCG GCTTCAGCAT CAACGAGTGC GTCCGCTTCA GCAAGTACTA 420

GCGCCTCAGC CTCAGCGTCA ACAAGTGCAT CGGCTTCAGC GTCAACGAGT GCGTCTGAGT 480

CAGCATCAAC GAGTGCGTCA GCCTCAGCAA GCACATCAGC TTCTGAATCT GCATCAACCA 540

GTGCGTCAGC CTCAGCATCG ACAAGCGCCT CAGCTTCAGC AAGTACCAGT GCGTCAGCCT 600

CAGCGTCGAC AAGTGCGTCG GCTTCAGCAA GTACCAGTGC GTCAGCCTCA GCAAGTACCA 660

GTGCGTCAGC CTCAGCGTCG ACAAGTGCGT CGGCCTCAAC CAGTGCATCT GAATCGGCAT 720

CAACCAGTGC GTCAGCCTCA GCAAGTACTA GTGCATCAGC TTCAGCATCA ACGAGTGCAT 780

CGGCTTCAGC ATCAACCAGT GCATCAGAGT CAGCAAGTAC CAGTGCGTCA GnTTCCGCAT 840

GCAACAAGTG CCTCGGCTTC AGCAAGTAC 869

(2) INFORMATION FOR SEQ ID NO: 370:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 750 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 370:

TCAACAAGTG CCTCAGCATC AGCATCAACG AGTGCGTCAG CCTCAGCAAG TACTAGTGCA 60

TCAGCATCAG CATCAACCAG TGCATCAGCC TCAGCAAGTA TCTCAGCGTC TGAATCGGCA 120

TCAACGAGTG CATCAGCATC AGCATCAACG AGTGCATCGG CTTCAGCGTC AACCAGTGCA 180

TCAGTCTCAG CAAGCACCAG TGCGTCGGCT TCAGCATCAA CGAGTGCCTC AGCCTCAGCA 240

AGTATCTCAG CGTCTGAATC GGCATCAACG AGTGCGTCAG CCTCAGCAAG TACTAGTGCA 300

TCGGCTTCAG CAAGCACCAG TGCGTCGGCT TCAGCATCAA CCAGTGCCTC AGCCTCAGCA 360

AGTATCTCAG CGTCTGAATC GGCATCAACG AGTGCGTCAG CCTCAGCAAG TACTAGTGCA 420

TCAGCATCAG CATCAACGAG TGCATCGGCT TCAGCAAGTA CCAGCGCCTC AGCTTCAGCA 480

AGCACCAGTG CGTCAGCCTC AGCAAGTACC AGCGCCTCAG CCTCAGCAAG CACCAGTGCC 540

TCAGCTTCAG CAAGTACCAG TGCGTCAGCC TCAGCGTCGA CAAGTGCGTC GGCTTCAGCA 600

AGTACCTCAG CGTCTGAATC AGCATCAACG AGTGCATCAG CTTCAGCATC AACAAGTGCT 660

TCAGCTTCAG CAAGTATCTC AGCGTCTGAA TCGGCATCAA CGAGTGCGTC CGCTTCAGCA 720

AGTACTAGCG CCTCAGCATC AGCGTCAACG 750

(2) INFORMATION FOR SEQ ID NO: 371:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 957 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 371:

CCGGAAAACA GCTCTGGCGC TTGGTCTTGC CCAGCGTATT GCTAGTGGTG ACGTGCCTGC 60

GGAAATGGCT AAGATGCGCG TGTTAGAACT TGATTTGATG AATGTCGTTG CAGGGACACG 120

CTTCCGTGGT GACTTTGAAG AACGCATGAA TAATATCATC AAGGATATTG AAGAAGATGG 180

CCAAGTCATC CTCTTTATCG ATGAACTCCA CACCATCATG GGTTCTGGTA GCGGGATTGA 240

TTCGACTCTG GATGCGGCCA ATATCTTGAA ACCAGCCTTG GCGCGTGGAA CTTTGAGAAC 300

GGTTGGTGCC ACTACTCAGG AAGAATATCA AAAACATATC GAAAAAGATG CGGCACTTTC 360

TCGTCGTTTC GCTAAAGTGA CGATTGAAGA ACCAAGTGTG GCAGATAGTA TGACTATTTT 420

ACAAGGTTTG AAGGCGACTT ATGAGAAACA TCACCGTGTA CAAATCACAG ATGAAGCGGT 480

TGAAACAGCG GTTAAGATGG CTCATCGTTA TTTAACCAGT CGTCACTTGC CAGACTCTGC 540

TATCGATCTC TTGGATGAGG CGGCAGCAAC AGTGCAAAAT AAGGCAAAGC ATGTAAAAGC 600

AGACGATTCA GATTTGAGTC CAGCTGACAA GGCCCTGATG GATGGCAAGT GGAAACAGGC 660

AGCCCAGCTA ATCGCAAAAG AAGAGGAAGT ACCTGTCTAC AAAGACTTGG TGACAGAGTC 720

TGATATTTTG ACCACCTTGA GTCGCTTGTC AGGAATCCCA GTTCAAAAAC TGACTCAAAC 780

GGATGCTAAG AAGTATTTAA ATCTTGAAGC AGAACTCCAT AAACGGGTTA TCGGTCAAGA 840

TCAAGCTGTT TCAAGCATTA GCCGTGCCAT TCGCCGCAAC CAGTCAGGGA TTCGCAGTCA 900

TAAGCGTCCG ATTGGTTCCT TTATGTTCCT AGGGCCTACA GGTGTCGGGG TATCCGA 957

(2) INFORMATION FOR SEQ ID NO: 372:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 807 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 372:

CAAAGCGCCT CAGCTTCAGC ATCAACAAGT GCGTCGGCTT CAGCATCAAC CAGTGCCTCG 60

GCTTCAGCGT CAACCAGTGC GTCACATTCA GCAAGTACCA GTGCTTCAGT CTCAGCATCA 120

ACAAGTGCTT CAGCCTCAGC ATCGACAAGT GCCTCGGCTT CAGCAAGCAC ATCAGCATCT 180

GAATCAGCGT CAACCAGTGC TTCGGCTTCA GCAAGTACCA GTGCTTCAGC TTCAGCATCA 240

ACCAGCGCCT CGGCCTCAGC AAGCACCTCA GCTTCTGAAT CGGCCTCAAC CAGCGCCTCG 300

GCCTCAGCAA GCACCTCAGC TTCTGAATCG GCCTCAACCA GCGCCTCAGC CTCAGCATCA 360

ACGAGTGCTT CGGCTTCAGC AAGCACAAGC GCCTCGGGTT CAGCATCAAC GAGTACGTCA 420

GCTTCAGCGT CAACCAGTGC TTCAGCCTCA GCATCAACAA GTGCGTCAGC CTCAGCAAGT 480

ATCTCAGCGT CTGAATCGGC ATCAACGAGT GCGTCTGAGT CAGCATCAAC GAGTACGTCA 540

GCCTCAGCAA GCACCTCAGC TTCTGAATCG GCCTCAACCA GTGCGTCAGC CTCAGCATCG 600

ACAAGCGCCT CAGCTTCAGC AAGTACCAGT GCTTCAGCCT CAGCGTCGAC AAGTGCGTCG 660

GCCTCAACCA GTGCATCTGA ATCGGCATCA ACCAGTGCGT CAGCCTCAGC AAGTACTAGT 720

GCATCGGCTT CAGCATCAAC CAGTGCCTCG GCTTCAGCGT CAACCAGTGC GTCAGCTTCA 780

GCAAGTACCA TGTGCTTCAT GTCTCAG 807

(2) INFORMATION FOR SEQ ID NO: 373:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1068 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 373:

CATCGGCTTC AGCATCAACG AGTGCGTCCG CTTCAGCAAG TACTACCGCC TCAGCCTCAG 60

CGTCAACAAG TGCATCGGCT TCAGCGTCAA CGAGTGCGTC TGAGTCAGCA TCAACGAGTG 120

CGTCACCTCA GCAAGCACAT CAGCTTCTGA ATCTGCATCA ACCAGTGCGT CACCTCAGCA 180

TCGACAAGCG CCTCAGCTTC AGCAAGTACC AGTGCGTCAC CTCAGCGTCG ACAAGTGCGT 240

CGGCTTCAGC AAGTACCAGT GCGTCAsCTC AGCAAGTACC AGTGCGTCAC CTCAGCGTCG 300

ACAAGTGCGT CGGCCTCAAC CAGTGCATCT GAATCGGCAT CAACCAGTGC GTCACCTCAG 360

CAAGTACTAG TGCATCAGCT TCAGCATCAA CGAGTGCATC GGCTTCAGCA TCAACCAGTG 420

CATCAGAGTC AGCAAGTACC AGTGCGTCAG cTTCCGCATC AACAAGTGCC TCGGCTTCAG 480

CAAGTACTAG CGCCTCAGCC TCAGCGTCAA CAAGTGCTTC AGCTTCCGCG TCAACCAGCG 540

CCTCGGCCTC AGCAAGTATC TCAGCGTCTG AATCGGCATC AACAAGTGCC TCGGCTTCAG 600

CATCAACGAG TGCATCAGTC TCAGCAAGCA CCAGTGCGTC GGCCTCAGCA AGCACCAGCG 660

CGTCTGAATC CGCATCAACC AGTGCCTCAG CTTCAGCAAG TACCTCAGCA TCTGAATCAG 720

CATCAACAAG TGCATCGGCT TCAGCAAGCA CAAGTGCTTC AGCCTCAGCA AGTATCTCAG 780

CGTCTGAATC GGCATCAACG AGTGCGTCCG CTTCAGCAAG TACTAGCGCC TCAGCATCAG 840

CGTCAACAAG TGCTTCGGCT TCAGCGTCAA CGAGTGCGTC TGAGTCAGCA TCAACGAGTA 900

CGTCAGCCTC AGCAAGCACA TCAGCTTCTG AATCTGCATC AACCAGTGCG TCAGCCTCAG 960

CATCGACAAG CGCCTCAGCT TCAGCAAGTA CCAGTGCGTC AGCCTCAGCA AGTACCAGTG 1020

CTTCAGCCTC AGCGTCGACA AGTGCGTCGG GCTCAACCAG TGCATCTG 1068

(2) INFORMATION FOR SEQ ID NO: 374:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 620 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 374:

CAGCATCAAC GAGTGCTTCA GTTTCAGCGT CAACCAGTGC CTCTGAATCA GCTTCAACAA 60

GTGCCTCGGC TTCAGCAAGC CCCAGTGCGT CGGCTTCAGC AAGTACTAGT GCATCGGCTT 120

CAGCATCGAC AAGTGCGTCT GAATCGGCAT CAACGAGTGC TTCGGCTTCA GCATCAACGA 180

GTGCGTCAGC CTCAGCAAGC ACATCAGCTT CTGAATCTGC ATcAACCAGT GCGTCCGyTT 240

CAGCGTCAAC CAGTGCGTCG GCTTCAGCGT CGACAAGTGC TTCGGCTTCA GCATCAACGA 300

GTGCGTCGGC CTCAGCAAGC GCAAGTACCT CAGCGTCAGC TTCCGCCTCA ACCAGTGCGT 360

CGGCTTCAGC AAGCACAAGT GCGTCAGCCT CAGCAAGTAT CTCAGCGTCT GAATCGGCAT 420

CAACGAGTGC GTCTGAGTCA GCATCAACGA GTACGTCAGC CTCAGCAAGC ACATCAGCTT 480

CTGAATCTGC ATCAACCAGT GCGTCAGCCT CAGCATCGAC AAGCGCCTCA GCTTCAGCAA 540

GTACCAGTGC TTCAGCCTCA GCGTCGACAA GTGCGTCGGC CTCAACCAGT GCATCTGAAT 600

CGGCATCAAC CAGTGCGTCA 620

(2) INFORMATION FOR SEQ ID NO: 375:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 720 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 375:

GTATTGGGGC GCCCCAACCT CTATGTGACT ACGGATTATT TCCTAGATTA CATGgGGATA 60

AACCATTTAG AAGAATTACC AGTGATTGAT GAGCTTGAGA TTCAAGCCCA AGAAAGCCAA 120

TTATTTGGTG AAAGGATAGA AGAAGATGAG AATCAATAAG TATATTGCCC ACGCAGGTGT 180

GGCCAGTAGG AGAAAAGCAG AAGAGCTGAT TAAGCAAGGC TTGGTGACGG TTAACGGCCA 240

AGTGGTGCGT GAACTAGCAA CCACTATCAA GTCAGGCGAC AAGGTCGAAG TTGAAGGTCA 300

ACCTATCTAC AACGAAGAAA AGGTCTACTA TCTGCTTAAC AAACCACGCG GTGTGATTTC 360

CAGTGTGACA GATGATAAGG GTCGCAAGAC GGTTGTCGAC CTCTTGCCCA ATGTCAAAGA 420

GCGTATTTAC CCTGTGGGTC GTTTGGACTG GGATACATCA GGTGTCTTGA TTTTGACCAA 480

TGATGGGGAC TTTACAGACG AGATGATTCA CCCTCGTAAT GAGATTGACA AGGTTTATGT 540

CGCGCGTGTT AAAGGTGTGG CCAATAAGGA CAATCTCCGC CCCTTGACCC GTGGTCTTGA 600

GATTGATGGT AAGAAAACCA AGCCATAATA TATAGGTTTT GTAGCCTCTA CACCATAAAT 660

ATTTGCTAAT AAAAATACTG TATTATTACC CTCTTAAGGT GCGAAATTAT TCAAGTTCTT 720

(2) INFORMATION FOR SEQ ID NO: 376:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 648 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 376:

CGCCATTTCC CATCGTACCG CCGAAAATCC CAGCGCCTCA GCCATCAAAT ATCCTATCAA 60

CGTTCTCAAA AAAAGTGACC GCTCTCTCAT CATGTTTCCA AGTGGTAGCC GCCACTCAAA 120

CGATGTCAAG GGGGGCGCAC ACTskATTGC CAAAATGGCC AAGGTCCGTA TCATGCCGGT 180

TACCTACACC GGTCCCATGA CTTTGAAGGG CTTGATTAGC CGTGAACGTG TCGATATGAA 240

CTTTGGAAAT CCAATCGATA TCTCAGATAT CAAGAAAATG AATGATGAAG GCATTGAAAC 300

AGTCGCCAAT CGTATTCAAA CAGAATTCCA ACGTCTGGAC GAAGAAACGA AACAATGGCA 360

CAATGATAAA AAACCAAATC CACTCTGGTG GTTTATCCGC ATCCCTGCCC TCATCCTTGC 420

TATTATCCTC GCTATCCTAA CCATCATCTT TAGCTTTATC GCAAGCTTCA TCTGGAACCC 480

AGATAAGAAA AGAGAAGAAC TTGCATAGAA GAAATGAACC TTGGCCAAAC AGCTAAGGTT 540

TTCATTTATA TAGTAGATTG GwACTAGAAT AGTACACCTC TACTTCTAAA ACATTTTTAG 600

AAATCGATTT GACTGTCCTG ATCGATTTGT CCTAATCTTA TTTCAATT 648

(2) INFORMATION FOR SEQ ID NO: 377:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 690 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 377:

GTGCATCGCT TTCAGCATCG ACAAGTGCGT CTGAATCGGC ATCAACGAGT GCTTCGGCTT 60

CAGCATCAAC GAGTGCGTCA GCTTCAGCAA GCACATCAGC TTCTGAATCT GCATCAACCA 120

GTGCGTCCGC TTCAGCGTCA ACCAGTGCGT CGGCTTCAGC GTCGACAAGT GCTTCGGCTT 180

CAGCATCAAC GAGTGCGTCG GCCTCAGCAA GCGCAAGTAC CTCAGCGTCA GCTTCCGCCT 240

CAACCAGTGC GTCCGCTTCA GCAAGCACAA GTGCGTCAGC CTCAGCAAGT ATCTCAGCGT 300

CTGAATCGGC ATCAACGAGT GCGTCGGCCT CAGCAAGCGC AAGTACCTCA GCGTCAGCTT 360

CCGCCTCAAC CAGTGCGTCG GCTTCAGCAA GCACAAGTGC GTCAGCCTCA GCAAGTATCT 420

CAGCGTCTGA ATCGGCATCA ACGAGTGCGT CTGAGTCAGC ATCAACGAGT ACGTCAGCCT 480

CAGCAAGCAC ATCAGCTTCT GAATCGGCAT CAACCAGTGC GTCAGCCTCA GCATCGACAA 540

GCGCCTCAGC TTCAGCAAGT ACCAGTGCTT CAGCCTCAGC GTCGACAAGT GCGTCGGCCT 600

CAACCAGTGC ATCTGAATCG GCATCAACCA GTGCGTCAGC CTCAGCAAGT ACTAGTGCAT 660

CAGCTTCAGC ATCAACGAGT GCATCGGCTT 690

(2) INFORMATION FOR SEQ ID NO: 378:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1003 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 378:

CGAGATTCTC TGGAGTTATG GATGTCGTTC CAATATGTGC ACGTTGGAAT GTTAGTGCTT 60

ATATGGGGGG AACAGAATCC TCTCTTGATT GAAGACAAGC TAGTCATTAG GCTGGTTTGT 120

CTTTTTGTCA ACTGTAGTGG GTTGATATAA TAGTATTAGT GAGTGGGATA AAAGTTTCAT 180

TTAGTTTATT CAGTACAAAT TTAACGGGTC AAGATTTATA TACTAGTGGT GTTTTTGGGG 240

CTGAGAGAAG TATCTTGATT TTATGTGTGG TTTTTATACT TACAGTTGTT CTGCTCCAAA 300

GAGCTTGTAG AGAAGAATTA GCTCATAAAG GAGATTGATT ATTTTGATAT CAAAAAAATG 360

CACAGGATAA CCTGATGCAT TTTTTTAGCG ACAATGCTTG CTACTTCCTT CTGTCGAATT 420

TAGACAATTT TAAACCCCAA TTATTCACCC CAAATCTAAA AACCATCCAG AATCCTTGCC 480

TTAGCTTAGA TCCTGGATGG TTTCTTTTTT CACCCAATGG GTGTTTTTTA CTAGACAAAA 540

AAGAGTTTCC CCTTTATGGT ATAAGTGTAG AAAAAAACAC AAAAAGAAAG GAAACTCACA 600

TGAACAGTTT ACCAAATCAT CACTTCCAAA ACAAGTCTTT TTACCAACTA TCTTTCGATG 660

GAGGTCATTT AACCCAGTAT GGTGGTCTTA TCTTTTTTCA GGAACTTTTT TCCCAGTTGA 720

AACTAAAAGA GCGGATTTCT AAGTATTTAG TAACGAATGA CCAACGCCGC TACTGTCGTT 780

ATTCGGATTC AGATATCCTT GTCCAGTTCC TCTTTCAACT GTTAACAGGT TATGGAACGG 840

ACTATGCTTG TAAAGAATTG TCAGCTGATG CCTACTTTCC AAAATTATTG GAAGGAGGGC 900

AGCTTGCTTC ACAGCCAACC TTATCCCGTT TTCTTTCCAG AACTGACGAG GAAACAGTCC 960

ATAGTTTGCG ATGCCTCAAC CTTGAATTGG TCGAATTCTT TTT 1003

(2) INFORMATION FOR SEQ ID NO: 379:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 738 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 379:

CCGATGATTC TGATTGGTTT GCTCTTTACT TTGCTGGGAA TTTTGAGGTA GATCTATGAT 60

TGAAATACTA ATTGTTTTAG CTATTATCCT ATCTCTTGCT TTGATTGTAT TGGTAACTAT 120

ACAACCCCGT CAAAATCAAC TATTTTCCAT GGATGCCACT AGTAATATTG GTAAACCAAG 180

CTACTGGCAG AGCAACACCT TGGTCAAGGT GCTCACTTTA TTGGTGAGTT TGGCTTTATT 240

TATTCTACTA TTAACCTTTA TGGTGATTAC TTATAAATAA AAGAAAACTT CAGATATTCA 300

CCTTTTGTGG ATTGGTCTGA AGTTTTCTTT TTTATACTCA ATGAAAATCA AAGAGCAAAC 360

TAGGAAGCTA GCCGCAckGC TCAAAACACC GTTTTGAGGT TGTAGATATA ACTGACGAGc 420

GACTCAAAAC ACCGTTTTGA GGTTGTAGAT ATAACTGACG AGcGACTCAA AACACCGTTT 480

TGAGGTTGTG GATAGAACTG ACGAGcGACT CAAAACACCG TTTTGAGGTT GTGGATAGAA 540

CTGACGAAGT CGcTCAAAAC ACCGTTTTGA GGTTGTGGAT AGAACTGACG AAtgctCAAA 600

ACACCGTTTT GAGGTTGTGG ATAGAACTGA CGAAGCgaaC ATATATACAG CAAGGCGACG 660

CTGACGTGGT TTGAAGAGTA TTACTGTCTA TATTTTTGGT AAAAATCAAC TTTTACTTGG 720

ATGAAGGTTT TTTTTTTT 738

(2) INFORMATION FOR SEQ ID NO: 380:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 695 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 380:

CCGTCTTATC AAAGAGGTTA ACAAAGGCAC CAAATTTCTC GATACGAACG ACTTTAGCAC 60

GGTAAACTTC ATCCACTTTG GCTTCACGAA CCAAACCAGC AATAATTTCT TTGGCACGGT 120

TAATAGCATC TTGGTCACTA GAGTAGATAG ACACATTTCC TTCTTCGTCT ATATCAATCT 180

TAACACCTGT TTCAGCGATA ATCTTGTCGA TGGTTTCTCC ACCCTTACCG ATGACAATCT 240

TAATCTTGTC CACATCAATC TTGATCGTAT CAATTTTCGG AGCAGTTGGA GCCAATTCTG 300

GACGAACTTC TGGAATGGTT GCTTCAATGA CATCAAGGAT TTCAAAACGC GCTTTCTTGG 360

CTTGAGCAAG AGCCTCCGTC AAGATTTCTG CAGTAATCCC TTGAATCTTG ATATCCATTT 420

GAAGGGCTGT AATCCCATCA CGAGTACCTG CAACCTTGAA GTCCATATCT CCAAAGTGAT 480

CTTCCAAACC TTGGATATCT GTCAATACTG TGTAGTTATT TCCATCTGAG ATAAGCCCCA 540

TAGCAATACC AGCTACTGGC GCCTTGATTG GCACACCACC AGCCATAAGG GCAAGAGTTC 600

CCGCACAGAT AGAAGCTTGA GATGAAGAAC CGTTTGATTC CAAAACTTCT GCTACTAGAC 660

GGATAGCGTA GGGGAATTCT TCCAAGCTTG GCAGG 695

(2) INFORMATION FOR SEQ ID NO: 381:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 691 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 381:

GACATCTTAT CTAAATACAT GCTAATATAT TTAGATACAA ACATTCCAAC TTGATAATTT 60

TCACTCATCT TTCATCATTC CTTATACAAC TATGCAGTAT AAATAGAATA GTTTTCTCAT 120

CAGAATGAGA CTATTTTAAT ATTAGATCCC CAATTATTCA CCCCAAATCT AAAAACCATC 180

CAGAATCCTT GCCTTAGCTT AGATCCTGGA TGGTTTCTTT TTTCACCCAA TGGGTGTTTT 240

TTACTAGACA AAAAAGAGTT TCCCCTTTAT GGTATAAGTG TAGAAAAAAA CACAAAAAGA 300

AAGGAAACTC ACATGAACAG TTTACCAAAT CATCACTTCC AAAACAAGTC TTTTTACCAA 360

CTATCTTTCG ATGGAGGTCA TTTAACCCAG TATGGTGGTC TTATCTTTTT TCAGGAACTT 420

TTTTCCCAGT TGAAACTAAA AGAGCGGATT TCTAAGTATT TAGTAACGAA TGACCAACGC 480

CGCTACTGTC GTTATTCGGA TTCAGATATC CTTGTCCAGT TCCTCTTTCA ACTGTTAACA 540

GGTTATGGAA CGGACTATGC TTGTAAAGAA TTGTCAGCTG ATGCCTACTT TCCAAAATTG 600

TTGGAAGGAG GGCAGCTTGc TTCACAGCCA ACCTTATCCC GwTTTCTTTC CAGAACTGAC 660

GAGGAAACAG TCCATAGTTT GCGATGCCTC A 691

(2) INFORMATION FOR SEQ ID NO: 382:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 750 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 382:

ATCTCTCTGC GTAATGGTCC TCAGATAACT CTGATGATGT GTGGCGATAT AGAACTGAGC 60

CAAGTTATGC CTAAAGGGCC TTAGGAATAG GAGCTTTCAC AAGCTTATCC AGATGATTAT 120

CTTTTACTCG TTATGGACAA TGCTATATGG CATAAATCAA GTACCTTAAA GATTCCGACT 180

AATATTGGCT TTGCATTTAT TCCTCCATAC ACACCAGAGA TGAACCCCAT TGAACAAGTG 240

TGGAAAGAGA TTCGTAAACG TGGATTTAAG AATAAAGCCT TTCGAACTTT GGAAGATGTC 300

ATACAAGGAC TGGAGAAGGA GGTGATAAAG TCCATCGTTA ATCGGAGACG GACTAGAATG 360

CTTTTTGAAA ACAGATGAGT ATAAAAAGAA AGTCCTCATT TCAATAGAAA TCACGACTTT 420

CTGATGAATT TATAGTAAAA TGAAATAAGA ACAGGATAGT CAAATCGATT TCTAACAATG 480

TTTTAGAAGC AGAGGTGTAC TATTCTAGTT TAAATCCACT ATATTTGGGG AGTGATAGAA 540

AAGCCCTTCA TCAGCCAATC TACTTGTTCA GGTGCGAGAG CTTTGACATC CTTTTCTGTA 600

CTGGACCAAG TCAGTTTTCC GTTCTCAAAG CGTTTATATA ATATCCAAAA TCCTTGACCA 660

TCCCAGTAAA GAACTTTAAA GCGGTCTTTA CGTCCACCAC AAAAGAGAAA GACTTGATCG 720

GAGAAAGGAT CCAATTCAAA GTGGGTTTGG 750

(2) INFORMATION FOR SEQ ID NO: 383:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 738 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 383:

TCAAATTCTT CGTGGTCCGC ATATCTnTCT TCGTACACGG CAGTCACTTG GTCTTTCACT 60

ACTCGAGTCG CAGCTTCACG GGCCAATTTC TCTTCTACTT GAACTGCCTT TTGGAGGTCA 120

CTGTTGTAGG CTGCAATGAT TTCAGCTTGC AATTCAGCAT CCACGTGAAG CAATTCCACT 180

TCTGCTTTTT CTTTACCGAC AGCAGCAACG ATTTCTTCTT GGAAGGCAAT CAATTCTTTG 240

ACAGCTTCGT GCCCTTTAAG GAGCGCTTCC AACATGATTT CTTCTGACAA TTCTTTGGCA 300

CCAGACTCTA CCATGTTGAT AGCGTGCTTG GTTCCAGCTA CTGTCAATTC AAGAAGAGAT 360

TGCTCTGCTT GTTCTTGACT TGGGTTGATG ATGATTTGGC CATCTACATA TCCCACTTGT 420

ACCCCAGCAA TTGGTCCGTC AAATGGAATA TCTGAAATAG ACAGTGCCAA AGATGAACCA 480

AACATAGCAG CCATTGGTGC AGATGCATTT TCATCATAAG AAAGCACTGT ATTGATGACT 540

TGGACTTCAT TACGGAAACC TTCCGCAAAC ATAGGACGAA TCGGACGGTC AATCAAACGC 600

GCTGTCAAGG TCGCATCTGT TGAAGGACGT CCTTCACGTT TCATAAAGCC ACCAGGAAAC 660

TTCCCAGCCG CATACATTTT TTCTTCGTAG TTGACTTGGA GTGGGAAGAA ATCCTCAGTT 720

GCCATTTTCT GGGGATCC 738

(2) INFORMATION FOR SEQ ID NO: 384:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 657 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 384:

CCCCCTATTT ACCGTGGACT AAAGTTGTAC AAGAAAAGTG CAAATAAGAA ATCTCCAGAT 60

TAGGAACTAT ATATGAGTTC TCTAGTCTGG AGATTTTTCA ATAGACTTCG TTATTGGGCG 120

GTTACTTTCG AAACTTTGAA AACTTCAAAA AACGGATTTT TATCGCTTTC AAATTCTTTT 180

GGGGTCAAAC TCAGTAACTT ATTCGCCTTG TAGACTTCAT GACGCTCAGG GTATACTTTC 240

AAGGTCCCAA ATAGCCAAGA ATCGTCAGCG ATATTATCTG AATCATCTCC TTCTTGTTCT 300

CCTTTAGTTC GCCTGAGGAC AGCCTTGACA CGCGCCAGAA TTCTCTAGGG CTAAAAGGCT 360

TGGTCAGGTA GTCATCAGCC CCTAATTCCA AGGCCAAAAC CTTATCAAAT TCATCACTTT 420

TCGCAGAAAC CATCATAATT GGAGTTTTGA CGCCTTTGGC TCTCAGCCGC TTACAAACTT 480

CCATGCCATC TAATTGTGGT AACATGATAT CAAGCAAGAT AAAATCAAAG GGTTCTGTTT 540

CTGCCAAAGC TAAGGCCTTC CGTCCATTTG TCACCAATTG AGTAGAAAAG CCTTCCTTAC 600

TTAAATGGTA GTCAAGCAAT TTCAGAATGT GTTCTTCATC ATCCACTAAT AAGACTT 657

(2) INFORMATION FOR SEQ ID NO: 385:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 586 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 385:

CCGCATCAGC ATCAACGAGT GCATCGGCTT CACGTCAACC AGTGCATCAG TCTCAGCAAG 60

CACCAGTGCG TCGGCTTCAG CATCAACGAG TGCCTCAGCC TCAGCAAGTA TCTCAGCGTC 120

TGAATCGGCA TCAACGAGTG CGTCAGCTCA GCAAGTACTA GTGCATCGGC TTCAGCAAGC 180

ACCAGTGCGT CGGCTTCAGC ATCAACCAGT GCCTCAGCCT CAGCAAGTAT CTCAGCGTCT 240

GAATCGGCAT CAACGAGTGC GTCACCTCAG CAAGTACTAG TGCATCAGCA TCAGCATCAA 300

CGAGTGCATC GGCTTCAGCA AGTACCAGCG CCTCAGCTTC AGCAAGCACC AGTGCGTCAC 360

CTCAGCAAGT ACCAGCGCCT CAGCCTCAGC AAGCACCAGT GCCTCAGCTT CAGCAAGTAC 420

CAGTGCGTCA CCTCAGCATC GACAAGTGCG TCGGCTTCAG CAAGTACCTC AGCGTCTGAA 480

TCAGCATCAA CGAGTGCGTC AGCTTCAGCA TCAACCAGTG CCTCAGCCTC AGCAAGTATC 540

AGTGCGTCAG CTTCAGCATC AACGAGTGCG TCAGCTGCAG CAAGTA 586

(2) INFORMATION FOR SEQ ID NO: 386:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 451 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 386:

CGTCGGCTTC AGCATCAACG AGTGCATCAG CTTCAGCATC AACAAGTGCT TCAGCTTCAG 60

CAAGTACCAG TGCGTCGGCT TCAGCATCAA CGAGTGCTTC AGTCTCAGCG TCAACCAGTG 120

CCTCTGAATC CGCATCAACA AGTGCCTCGG CTTCAGCAAG CACCAGTGCT TCGGCTTCAG 180

CGTCAACGAG TGCGTCTGAG TCAGCATCAA CGAGTGCGTC ACCTCAGCAA GCACATCAGC 240

TTCTGAATCT GCATCAACCA GTGCGTCAGC TTCCGCATCA ACAAGCGCCT CGGCCTCAGC 300

AAGTACAAGT GCTTCAGCCT CAGCATCAAC CAGTGCATCA GCTTCAGCCT CAACAAGTGC 360

TTCAGCCTCA GCGTCAACCA GTGCCTCGGC TTCAGCAAGT ACCAGTGCGT CAGTTcAGCA 420

AGCACAAGTG CGTCAATTTA GCATCAACCA G 451

(2) INFORMATION FOR SEQ ID NO: 387:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 425 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 387:

TCTCAGCAAG CACCATTGCG TCGGCTTCAT CAAGCACCAG CGCGTTTGAA TCCGCATCAA 60

CCAGTGCTTC AGCTTCAGCC AAGTTACCTC AGCATCTGAA TCAGCATCAA CAAGTGCATC 120

GGCTTCAGCA AGCACAAGTG CTTCAGCtCA GCAAGTATCT CAGCGTCTGA ATCGGCATCA 180

ACGAGTGCGT CCGCTTCAGC AAGTACTAGC GCCTCAGCAT CAGCGTCAAC AAGTGCTTCG 240

GCTTCAGCGT CAACGAGTGC GTCTGAGTCA GCATCAACGA GTACGTCAGC CTCAGCAAGC 300

ACATCAGCTT CTGAATCTGC ATCAACCAGT GCGTCAGCCT CAGCATCGAC AAGCGCCTCA 360

GCTTCAGCAA GTACCAGTGC GTCAGCCTCA GCAAGTACCA GTGCTTCAGC CTCAGCGTCG 420

ACAAG 425

(2) INFORMATION FOR SEQ ID NO: 388:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 572 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 388:

AGAGGATCCC CGGATCCTCA GTCGCTGAGA TAACTCCTTT GGGCTTGTTC ATCATGTAGT 60

AGACAAACTC TTCATACTCC AACACTTGCC CATTTTATGC GAATCTCATC TATTTTTTCT 120

TTTTTTTGCA ATTTAGCTGA TTTTTCTTTT TTACCATTTA CAGTCACGCG CCCAGCCTTG 180

AGCAAGTTTT TGACCTCAGT CCGACTTCCC ACCGCACAGG CAACTAAAAA TTTATCTAAT 240

CTCATAGAAC TATTATATCA TATCAAAAGG AGGCTAGTAC AATGACCAAC CTCCTTTTCG 300

TTTCATACTC TTCAAAAATC TCTTCAAACC GCGTCAACGT CGCCTTGCCG TATATATGTT 360

ACTGACTTCG TCAGTTCTAT CTGCAACCTC AAAACAGTGT TTTGAGCTGA CTTCGTCAGT 420

TCTATCTGCA ACCTCAAAGC AGTGCTTTGA GCATCCTGCG GCTAGTTTCC kAGTkTGCTC 480

TTTGATTTwC ATTGAGTATC AGATTTAGGA AATTAACTTC CTCGkCTCCA AAAAAkAGCT 540

AAAACAATCA AGGCTCCTAA AATCGCTGGG AT 572

(2) INFORMATION FOR SEQ ID NO: 389:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 505 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 389:

CAACAAGTGC CTCGGCTTCA GCATGCACAA GTGCTTCAGC TTCAGCATGT ACCTGAGCGT 60

CTGAATCAGC ATCAACGTGT GCGTCCGCTT CAGCATGTAC TGCTGCCTCA GCATCAGCGT 120

CAAcAwGTGC TTCGGCTTCA GCGTCAACGA GTGCGTCTGA GTCAGCATCA ACGAGTACGT 180

CAGCCTCAGC AAGCACATCA GCTTCTGAAT CTGCATCAAC CAGTGCGTCA GCCTCAGCAT 240

CGACAAGCGC CTCAGCTTCA GCAAGTACCA GTGCGTCAGC CTCAGCAAGT ACCAGTGCTT 300

CAGCCTCAGC GTCGACAAGT GCGTCGGCCT CAACCAGTGC ATCTGAATCG GCATCAACCA 360

GTGCGTCAGC CTCAGCAAGT ACTAGCGCCT CAGCCTCAGC ATCAACGAGT GCGTCCGCTT 420

CAGCAAGTAC TAGTGCATCA GCATCAGCAT CAACGAGTGC ATCGGCTTCA GCAAGTACCA 480

GCGCCTCAGC TTCAGCAAGC ACCGG 505

(2) INFORMATION FOR SEQ ID NO: 390:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 447 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 390:

GCTAAGACTA CCTCATTAGG GGCATAGGCT GCTAAAATAA CTGCAGCTGT GGTTAATGAC 60

AATACTGTAC TTTTTTTCAT TTTAATTCCT TACATATTTA TATAACTTCC AATAGATAAT 120

AAACTTTAAC TTTGCTAGCC TTTGTTATAA AAAGTTTTAC TAAGTATTAT CTAGGAAATA 180

GAGTAGTACA TTTATATATA ATTGTTATCT CTCTATAAAA ACAGTATATC ATTTAAAAAA 240

ATTTAAGTCA AAAAAATTAA CATTAGTTAA TTTATTTTTT AGCACACATT AAAAAATAAG 300

ATTAGTACTC AATGAAAATC AAAGAGCAAA CTAGGAAACT AGCCGCAGAT TGCTCAAAAC 360

AGTGTTTTGA GGTTGTAGAT GGAATGACGT AGTCAGCTCA AAACACTGTT TTGAAGTTGT 420

GGATAGAACT GACGAAGTCG GTACCGA 447

(2) INFORMATION FOR SEQ ID NO: 391:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 572 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 391:

AGCACTTGTC GTTGAATTCT ACAACAAAAT GTTGTAATAT TTTATTGAAT AAGATAGGCC 60

TTGATATTAA GCACTTTGGG ACGTTCTCCC TTAGTGCTTT TTTGATTTCT CTTAGTATCC 120

AGCTATAATC GTTGAGACAT AACTAGACCG ATATAGTCCA AAGTGATATA GTAAAATGAA 180

CCAAAAATAG TACACAATGT GGTATAATCC TTTTATGGCA TATTCAATAG ATTTTCGTAA 240

AAAAGTTCTC TCTTATTGTG AGCGAACAGG TAGTATAACA GAAGCATCAC ACGTTTTCCA 300

AATCTCACGT AATACCATTT ATGGCTGGTT AAAGCTAAAA GAGAAAACAG GAGAGCTAAA 360

CCACCAAGTA TAGTGTATTG AATCTATAAC AGTACACCTT GGCTGCTAAA ATATTTCTAT 420

AAATTAATTT GACTTTCCTG ATAGAGATGT TCACATCTTA TTTCAAACTA CTATATAAGT 480

TCTATAATCT CTTTATAAGA TTTGCCCATC AGACAAAATA GAACGATTTG AAGGCGTTTA 540

TGATATTTAG CTGTACGAGA GTCTTTTAAA AG 572

MISSING UPON TIME OF PUBLICATION

DENMARK

The applicant hereby requests that, until the application has been laid open to public inspection (by the Danish Patent Office), or has been finally decided upon by the Danish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Danish Patent Office not later than at the time when the application is made available to the public under Sections 22 and 33(3) of the Danish Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Danish Patent Office or any person approved by the applicant in the individual case.

SWEDEN

The applicant hereby requests that, until the application has been laid open to public inspection (by the Swedish Patent Office), or has been finally decided upon by the Swedish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the International Bureau before the expiration of 16 months from the priority date (preferably on the Form PUT/RO/134 reproduced in annex Z of Volume I of the PCT Applicant's Guide). If such a request has been filed by the applicant, any request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Swedish Patent Office or any person approved by the applicant in the individual case.

UNITED KINGDOM

The applicant hereby requests that the furnishing of a sample of a microorganism shall only be made available to an expert. The request to this effect must be filed by the applicant with the International Bureau before the completion of the technical preparations for the International publication of the application.

NETHERLANDS

The applicant hereby requests that until the date of a grant of a Netherlands patent or until the date on which the application is refused or withdrawn or lapse, the microorganism shall be made available as provided in Rule 31F(1) of the Patent Rules only by the issue of a sample to an expert. The request to this effect must be furnished by the applicant with the Netherlands Industrial Property Office before the date on which the application is made available to the public under Section 22C or Section 25 of the Patents Act of the Kingdom of the Netherlands, whichever two dates occurs earlier.

SINGAPORE

The applicant hereby requests that the furnishing of a sample of a microorganism shall only be made available to an expert. The request to this effect must be filed by the applicant with the International Bureau before the completion of the technical preparations for international publication of the application.

NORWAY

The applicant hereby requests that, until the application has been laid open to public inspection (by the Norwegian Patent Office), or has been finally decided upon by the Norwegian Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Norwegian Patent Office not later than at the time when the application is made available to the public under Sections 22 and 33(3) of the Norwegians Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Norwegian Patent Office or any person approved by the applicant in the individual case.

AUSTRALIA

The applicant hereby gives notice that the furnishing of a sample of a microorganism shall only be effected prior to the grant of a patent, or prior to the lapsing, refusal or withdrawal of the application, to a person who is a skilled addressee without an interest in the invention (Regulation 3.25(3) of the Australian Patents Regulations).

FINLAND

The applicant hereby requests that, until the application has been laid open to public inspection (by the National Board of Patents and Registration), or has been finally decided upon by the National Board of Patents and Registration without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art.

ICELAND

The applicant hereby requests that, until the application has been laid open to public inspection (by the Icelandic Patent Office), or has been finally decided upon by the Icelandic Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected in the art.

What Is Claimed Is:

1. Computer readable medium having recorded thereon the nucleotide sequence depicted in SEQ ID NOS:1-391, a representative fragment thereof or a nucleotide sequence at least 95% identical to a nucleotide sequence depicted in SEQ ID NOS:1-391.

2. Computer readable medium having recorded thereon any one of the fragments of SEQ ID NOS:1-391 depicted in Tables 2 and 3 or a degenerate variant thereof.

3. The computer readable medium of claim 1, wherein said medium is selected from the group consisting of a floppy disc, a hard disc, random access memory (RAM), read only memory (ROM), and CD-ROM.

4. The computer readable medium of claim 3, wherein said medium is selected from the group consisting of a floppy disc, a hard disc, random access memory (RAM), read only memory (ROM), and CD-ROM.

5. A computer-based system for identifying fragments of the *Streptococcus pneumoniae* genome of commercial importance comprising the following elements:

a) a data storage means comprising the nucleotide sequence of SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence at least 95% identical to a nucleotide sequence of SEQ ID NOS:1-391;

b) search means for comparing a target sequence to the nucleotide sequence of the data storage means of step (a) to identify homologous sequence(s), and

c) retrieval means for obtaining said homologous sequence(s) of step (b).

6. A method for identifying commercially important nucleic acid fragments of the *Streptococcus pneumoniae* genome comprising the step of comparing a database comprising the nucleotide sequences depicted in SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence at least 95% identical to a nucleotide sequence of SEQ ID NOS:1-391 with a target sequence to obtain a nucleic acid molecule comprised of a complementary nucleotide sequence to said target sequence, wherein said target sequence is not randomly selected.

7. A method for identifying an expression modulating fragment of *Streptococcus pneumoniae* genome comprising the step of comparing a database comprising the nucleotide sequences depicted in SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence at least 95% identical to the nucleotide sequence of SEQ ID NOS:1-391 with a target sequence to obtain a nucleic acid molecule comprised of a complementary nucleotide sequence to said target sequence, wherein said target sequence comprises sequences known to regulate gene expression.

8. An isolated protein-encoding nucleic acid fragment of the *Streptococcus pneumoniae* genome, wherein said fragment consists of the nucleotide sequence of any one of the fragments of SEQ ID NOS:1-391 depicted in Tables 2 and 3, or a degenerate variant thereof.

9. A vector comprising any one of the fragments of the *Streptococcus pneumoniae* genome SEQ ID NOS:1-391 depicted in Tables 2 and 3 or a degenerate variant thereof.

10. An isolated fragment of the *Streptococcus pneumoniae* genome, wherein said fragment modulates the expression of an operably linked open reading frame, wherein said fragment consists of the nucleotide sequence from about 10 to 200 bases in length which is 5' to any one of the open reading frames depicted in Tables 2 and 3 or a degenerate variant thereof.

11. A vector comprising any one of the fragments of the *Streptococcus pneumoniae* genome of claim 8.

12. An organism which has been altered to contain any one of the fragments of the *Streptococcus pneumoniae* genome of claim 8.

13. An organism which has been altered to contain any one of the fragments of the *Streptococcus pneumoniae* genome of claim 10.

14. A method for regulating the expression of a nucleic acid molecule comprising the step of covalently attaching to said nucleic acid molecule a nucleic acid molecule consisting of the nucleotide sequence from about 10 to 100 bases 5' to any one of the fragments of the *Streptococcus pneumoniae* genome depicted in SEQ ID NOS:1-391 and Tables 2 and 3 or a degenerate variant thereof.

15. An isolated nucleic acid molecule encoding a homolog of any of the fragments of the *Streptococcus pneumoniae* genome of SEQ ID NOS:1-391 and Tables 2 and 3, wherein said nucleic acid molecule is produced by a process comprising steps of:

a) screening a genomic DNA library using as a probe a target sequence defined by any of SEQ ID NOS:1-391 and Tables 2 and 3, including fragments thereof;

b) identifying members of said library which contain sequences that hybridize to said target sequence; and

c) isolating the nucleic acid molecules from said members identified in step (b).

16. An isolated DNA molecule encoding a homolog of any one of the fragments of the *Streptococcus pneumoniae* genome of SEQ ID NOS:1-391 and Tables 2 and 3, wherein said nucleic acid molecule is produced a process comprising steps of:

a) isolating mRNA, DNA, or cDNA produced from an organism;

b) amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from said fragment of said *Streptococcus pneumoniae* genome to prime said amplification;

c) isolating said amplified sequences produced in step (b).

17. An isolated polypeptide encoded by any of the fragments of the *Streptococcus pneumoniae* genome of SEQ ID NOS:1-391 and depicted in Table 2 and 3 or by a degenerate variant of said fragments.

18. An isolated polynucleotide molecule encoding any one of the polypeptides of claim 17.

19. An antibody which selectively binds to any one of the polypeptides of claim 17.

20. A method for producing a polypeptide in a host cell comprising the steps of:

a) incubating a host containing a heterologous nucleic acid molecule whose nucleotide sequence consists of any one of the fragments of the *Streptococcus pneumoniae* genome of SEQ ID NOS:1-391 and depicted in Tables 2 and 3, under conditions where said heterologous nucleic acid molecule is expressed to produce said protein, and

b) isolating said protein.

Figure 1

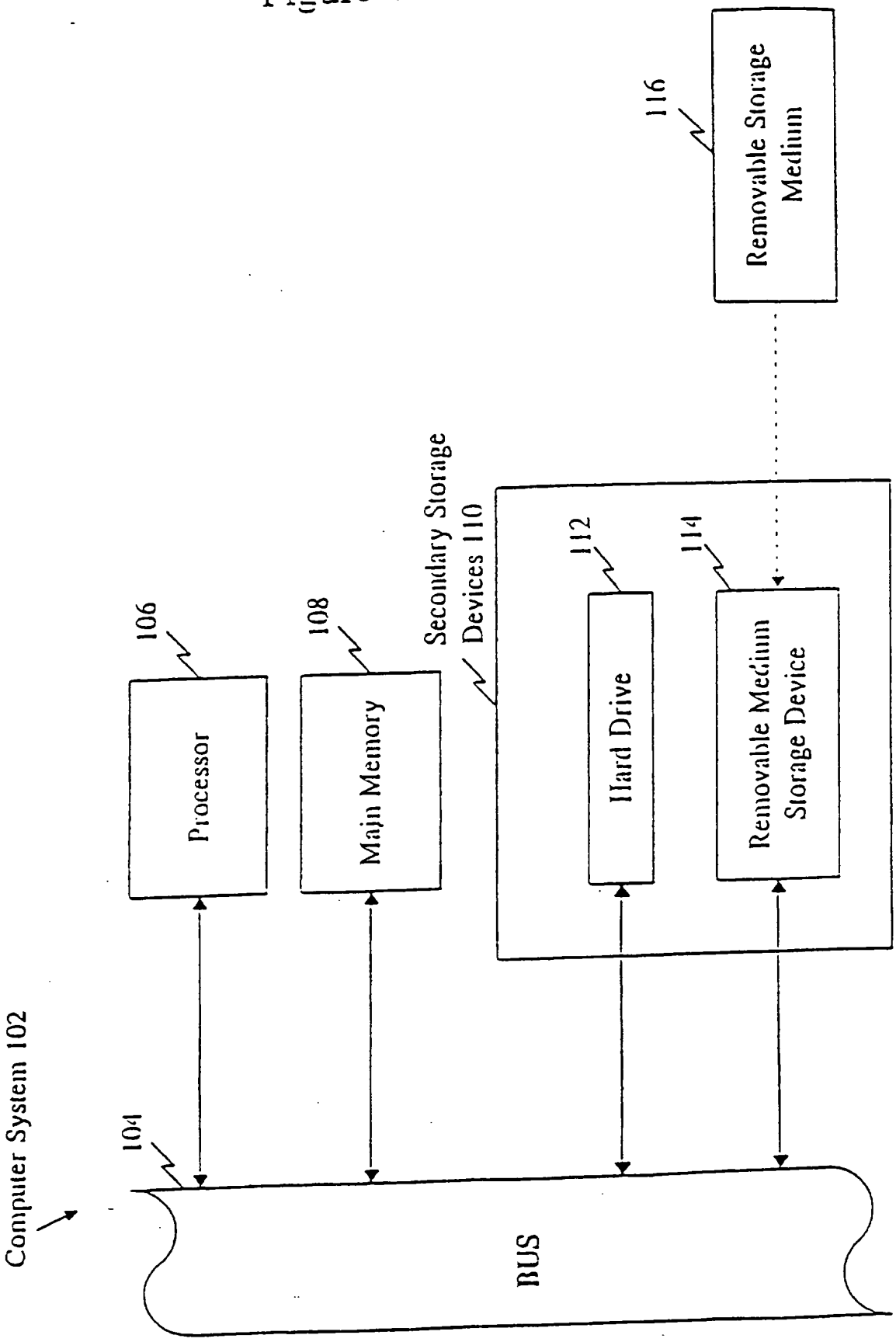
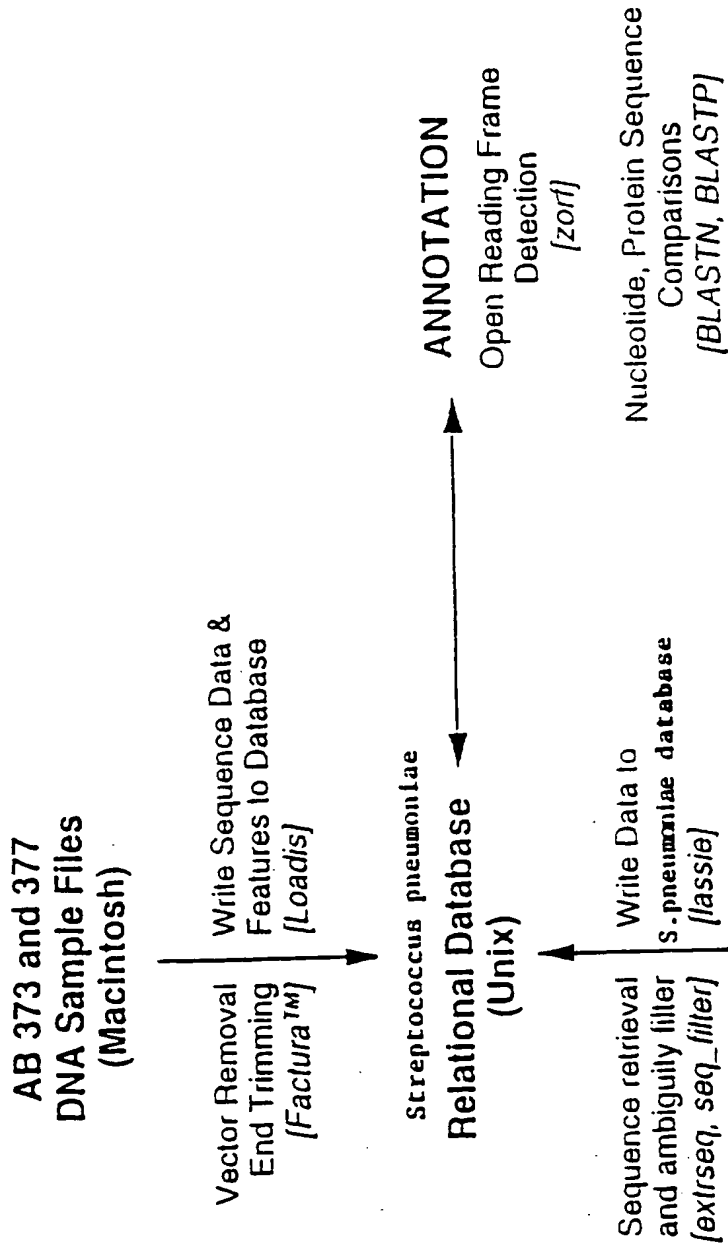

Computer System 102
104
BUS
106
Processor
108
Main Memory
Secondary Storage Devices 110
112
Hard Drive
114
Removable Medium Storage Device
116
Removable Storage Medium

Figure 2


AB 373 and 377
DNA Sample Files
(Macintosh)
Vector Removal
End Trimming
[Factura™]
Write Sequence Data &
Features to Database
[Loadis]
Streptococcus pneumoniae
Relational Database
(Unix)
ANNOTATION
Open Reading Frame
Detection
[zorf]
Nucleotide, Protein Sequence
Comparisons
[BLASTN, BLASTP]
Sequence retrieval
and ambiguity filter
[extrseq, seq_filter]
Write Data to
S.pneumoniae database
[lassie]
ASSEMBLY
Rapid Assembly
and Ordering of
Thousands of Sequences
[TIGR Assembler]
[asm_align]

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/31, C07K 14/315, 16/12, C12Q 1/68 — A3

(11) International Publication Number: **WO 98/18931**

(43) International Publication Date: 7 May 1998 (07.05.98)

(21) International Application Number: PCT/US97/19588

(22) International Filing Date: 30 October 1997 (30.10.97)

(30) Priority Data:
60/029,960 31 October 1996 (31.10.96) US

(71) Applicant *(for all designated States except US)*: HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: KUNSCH, Charles, A. [US/US]; 2398B Dunwoody Crossing, Atlanta, GA 30338 (US). CHOI, Gil, H. [KR/US]; 11429 Potomac Oaks Drive, Rockville, MD 20850 (US). DILLON, Patrick, J. [US/US]; 1055 Snipe Court, Carlsbad, CA 92009 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). BARASH, Steven, C. [US/US]; 582 College Parkway #303, Rockville, MD 20850 (US). FANNON, Michael [US/US]; 13501 Rippling Brook Drive, Silver Spring, MD 20850 (US). DOUGHERTY, Brian, A. [US/US]; 708 Meadow Field Court, Mount Airy, MD 21771 (US).

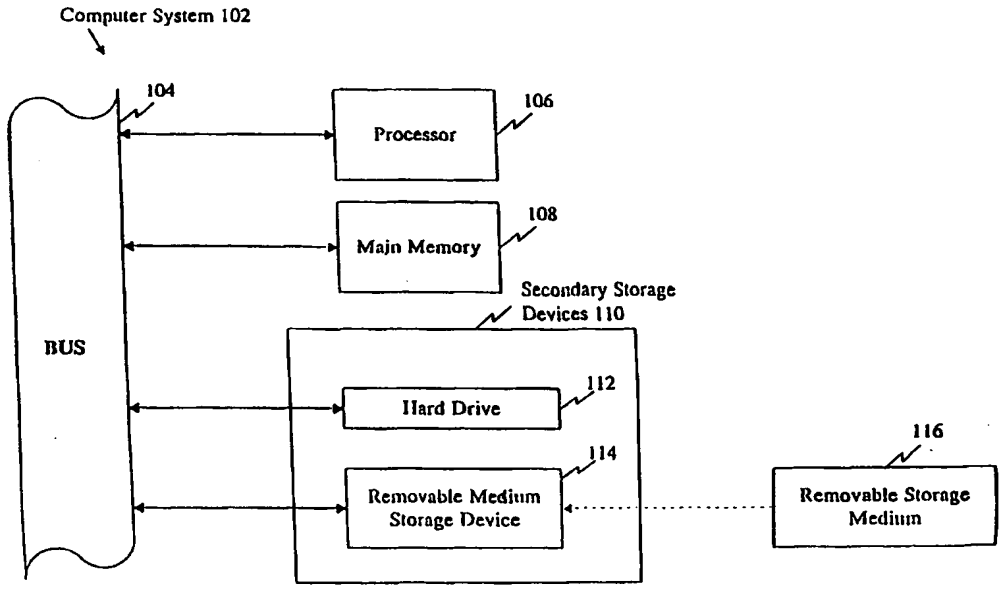
(74) Agents: BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
20 August 1998 (20.08.98)

(54) Title: *STREPTOCOCCUS PNEUMONIAE* POLYNUCLEOTIDES AND SEQUENCES



(57) Abstract

The present invention provides polynucleotide sequences of the genome of *Streptococcus pneumoniae*, polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynucleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

In. .ational Application No
PCT/US 97/19588

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/31 C07K14/315 C07K16/12 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 96 33276 A (HUMAN GENOME SCIENCES INC ;UNIV JOHNS HOPKINS (US)) 24 October 1996 see claims 1-7 ---	1-7
A	ALTSCHUL S F ET AL: "BASIL LOCAL ALIGNMENT SEARCH TOOL" JOURNAL OF MOLECULAR BIOLOGY, vol. 215, 1990, pages 403-410, XP000604562 cited in the application see the whole document --- -/--	1-7

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

4

Date of the actual completion of the international search	Date of mailing of the international search report
27 March 1998	0 8. 07. 98
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer HORNIG H.

Form PCT/ISA/210 (second sheet) (July 1992)

Int. tional Application No
PCT/US 97/19588

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	W.R. PEARSON AND D.J. LIPMAN: "Improved tools for biological sequence comparison" PROC. NATL. ACAD. SCI., vol. 85, April 1988, NATL. ACAD. SCI.,WASHINGTON,DC,US;, pages 2444-2448, XP002060460 cited in the application see the whole document	1-7
A	WO 95 06732 A (UNIV ROCKEFELLER ;MASURE H ROBERT (US); PEARCE BARBARA J (US); TUO) 9 March 1995 see the whole document	1-7
A	WO 95 31548 A (UAB RESEARCH FOUNDATION ;YOTHER JANET (US); DILLARD JOSEPH P (US)) 23 November 1995 see the whole document	1-7
A	WO 95 14712 A (RES CORP TECHNOLOGIES INC) 1 June 1995 see the whole document.	1-7
A	WO 96 05859 A (AMERICAN CYANAMID CO) 29 February 1996 see the whole document	1-7
A	WO 93 10238 A (US HEALTH) 27 May 1993 see the whole document	1-7
A	EP 0 687 688 A (UNIV OVIEDO ;UNIV LEICESTER (GB)) 20 December 1995 see the whole document	1-7
A	EP 0 622 081 A (UAB RESEARCH FOUNDATION) 2 November 1994 see the whole document	1-7

INTERNATIONAL SEARCH REPORT

...ternational application No.
PCT/US 97/ 19588

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
Remark: Although claims 1-4 could be, at least partially be considered as a mere presentation of information Rule 39.1(v), and claims 5-7 at least partially as a computer program, Rule 39.1(vi)PCT, the search has been carried out as far as possible in our systematic documentation.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see continuation-sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-7

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-7

Computer readable medium having recorded thereon the nucleotide sequence depicted in SEQ ID nos. 1-391, a representative fragment thereof or a nucleotide sequence at least 95% identical to a nucleotide sequence depicted in SEQ ID nos. 1-391; a computer-based system for identifying fragments of the Streptococcus pneumoniae genome of commercial importance comprising: a) a data storage means comprising said nucleotide sequence(s); b) search means for comparing a target sequence to the nucleotide sequence of the data storage means of step (a) to identify homologous sequence(s), and c) retrieval means for obtaining said homologous sequence(s) of step (b); a method for identifying commercially important nucleic acid fragments of the Streptococcus pneumoniae genome comprising the step of comparing a database comprising said nucleotide sequence(s) with a target sequence to obtain a nucleic acid molecule comprised of a complementary nucleotide sequence to said target sequence, wherein said target sequence is not randomly selected; a method for identifying an expression modulating fragments of the Streptococcus pneumoniae genome comprising the step of comparing a database comprising said nucleotide sequence(s) with a target sequence to obtain a nucleic acid molecule comprised of a complementary nucleotide sequence to said target sequence, wherein said target sequence comprises sequences known to regulate gene expression;

2. Claims: (8-20) partially

An isolated protein-encoded nucleic acid fragment of the Streptococcus pneumoniae genome, wherein said fragment consists of the nucleotide sequence of the fragment of SEQ ID no.1 depicted in Tables 2 and 3, or a degenerate variant thereof; a vector comprising the fragment of the Streptococcus pneumoniae genome SEQ ID no.1; an isolated fragment of the Streptococcus pneumoniae genome, wherein said fragment modulates the expression of an operably linked open reading frame, wherein said fragment consists of the nucleotide sequence from about 10 to 200 bases in length which is 5' to any one of the open reading frame of SEQ ID no.1 depicted in Tables 2 and 3 or a degenerate variant thereof; a method for regulating the expression of a nucleic acid molecule comprising the step of covalently attaching to said nucleic acid molecule a nucleic acid molecule consisting of the nucleotide sequence from about 10 to 100 bases 5' to any one of the open reading frame of SEQ ID no.1 and Tables 2 and 3 or a degenerate variant thereof; an isolated nucleic acid molecule encoding a homolog of SEQ ID no.1; an isolated polypeptide encoded by SEQ ID no.1 and depicted in Table 2 and 3; an antibody which selectively binds to any one of said polypeptides, a method for producing a polypeptide in a host cell comprising a) incubating a host containing a heterologous nucleic acid

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 97/09588

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9633276 A	24-10-96	AU 5552396 A	07-11-96
		EP 0821737 A	04-02-98
WO 9506732 A	09-03-95	AU 7680994 A	22-03-95
		CA 2170726 A	09-03-95
		EP 0721506 A	17-07-96
		FI 960977 A	30-04-96
		JP 9504686 T	13-05-97
		NO 960839 A	19-04-96
WO 9531548 A	23-11-95	AU 2638595 A	05-12-95
		EP 0804582 A	05-11-97
WO 9514712 A	01-06-95	US 5474905 A	12-12-95
WO 9605859 A	29-02-96	US 5565204 A	15-10-96
		AU 3363695 A	14-03-96
		CA 2198251 A	29-02-96
		EP 0778781 A	18-06-97
		JP 10504717 T	12-05-98
WO 9310238 A	27-05-93	AU 3065892 A	15-06-93
EP 0687688 A	20-12-95	ES 2075803 A	01-10-95
		ES 2088820 A	16-09-96
		WO 9516711 A	22-06-95
EP 0622081 A	02-11-94	AU 682018 B	18-09-97
		AU 5769694 A	27-10-94
		CA 2116261 A	21-10-94
		FI 941695 A	21-10-94
		JP 7126291 A	16-05-95
		NO 941420 A	21-10-94
		US 5679768 A	21-10-97
		ZA 9401584 A	12-10-94

Attach #11 08/840,589

molecule whose nucleotide sequence consists of SEQ ID no.1 and depicted in Table 2 and 3, under conditions where said heterologous nucleic acid molecule is expressed to produce said protein, and b) isolating said protein;

3-392. Claims:(8-20) partially

Idem as subject 2 but limited to each of the sequences of SEQ ID no. 2 to 391;

For the sake of conciseness, the second subject matter is explicitly defined, the other subject matters are defined by analogy hereto.